DOCUMENT RESUME

ED 181 052 TM 009 954

TITLE	Proceedings of the Invitational Conference on Testing Problems (New York, New York, October 28, 1961).
INSTITUTION	Educational Testing Service, Princeton, N.J.
PUB DATE	28 Oct 61
NOTE	107p.
EDRS PRICE	MF01/PC05 Plus Postage.
DESCRIPTORS	*Achievement Tests: Advanced Placement Programs: *College Credits: College Entrance Examinations: College Instruction: Composition (Literary): Conference Reports: Educational Quality: Educational Television: *Equivalency Tests: Evaluation: *Factor Analysis: Higher Education: Institutional Research: Measurement Techniques: Senior High Schools: *Testing Problems: Testing Programs: Test Validity: Verbal Ability
IDENTIFIERS	Antioch College OH: Content Validity: Writing Evaluation

ABSTRACT

Three themes were addressed at the conference: (1) implications of factor analysis for achievement testing: (2) use of achievement tests in awarding course credits: and (3) extended conceptions of evaluation in higher education. The speeches were entitled: Factors of Verbal Achievement, by John B. Carroll: Schools of Thought in Judging Excellence of English Themes, by John W. French: Factor Analysis of Relevance Judgments: an Approach to Content Validity, by Ledyard R Tucker: Credit and the Advanced Placement Program, by Jack N. Arbolino: The Examining Program at Antioch College, by Ruth D. Churchill: Examinations and Instruction by Television, by Clifford G. Erickson: Credit and Waiver Examinations at Michigan State University, by Willard G. Warrington: Measuring the Quality of a College or University, by Dewey B. Stuit: and the Nature and Use of Institutional Research, by James I Doi. The luncheon address, Management of the Learning Environment, was delivered by John E. Ivey, Jr. A memorial speech in honor of Irving D. Lorge was given by Robert I. Thorndike. (MH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED181052

INVITATIONAL CONFERENCE ON TESTING PROBLEMS

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Educational Testing Service
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

1961

EDUCATIONAL TESTING SERVICE

Copyright, 1962, Educational Testing Service

Princeton, N. J.

Library of Congress Catalogue Number: 47-11220

Printed in the United States of America

Invitational Conference on Testing Problems

October 28, 1961
Hotel Roosevelt
New York City

PAUL L. DRESSEL
Chairman

Educational Testing Service
Princeton, New Jersey
Los Angeles, California

ETS
Board of Trustees
1961-62

Caryl P. Haskins, *Chairman*

John S. Allen
George F. Baughman
William G. Carr
John H. Fischer
John W. Gardner
Mason W. Gross
Stuart Hedden
Philip J. Hickey
Arnold E. Joyal
Katharine E. McBride
John D. Millett
Edward S. Noyes
B. Alden Thresher
Logan Wilson

ETS Officers

Henry Chauncey, *President*
William W. Turnbull, *Executive Vice President*
Henry S. Dyer, *Vice President for College Board Programs*
Robert L. Ebel, *Vice President for General Programs*
G. Dykeman Sterling,
Vice President for Administration and Treasurer
Catherine G. Sharp, *Secretary*
Paul G. Haskell, *Assistant Treasurer*

Foreword

THE 1961 Invitational Conference on Testing Problems maintained the high standard set by its predecessors. For this we are particularly indebted to Dr. Paul L. Dressel of Michigan State University, who planned and organized it. We are also most grateful to the distinguished speakers who contributed their time and effort in making this such an outstanding Conference. We are particularly grateful for the stimulating luncheon address on "Management of the Learning Environment," by Dr. John E. Ivey, Jr.

These *Proceedings* will, I am sure, be of great interest to all educators who are concerned with new developments in the field of testing.

Henry Chauncey

President

Preface

THE ANNUAL Invitational Conference on Testing Problems is one of the major events each year for those interested in the fields of measurement and evaluation. In sponsoring this conference the Educational Testing Service contributes most effectively to the spread of new ideas in the field and to the development of bonds of common concern among those in the testing field.

The wide range of interests and convictions reflected in the persons attending the conference is one of its strengths. At the same time, this heterogeneity poses a problem for the program planner who would appeal to all interests. In planning the program this year, three major problem areas of widespread interest were defined. The first problem was that of the contribution of factor analytic studies to the understanding and improvement of achievement measurements. The papers successively presented by John B. Carroll, John W. French, and Ledyard R Tucker admirably supplement each other in pointing to factors—both subjective and objective—which underlie the problems of theme reading and achievement testing.

The second area of concern covered was that of the use of achievement tests in award of course credit. Although tests have long been used in European education for granting recognition and degrees, and have even been used extensively in this country at various periods, the practice is still regarded as novel and debatable. The four papers presented by Jack N. Arbolino, Ruth D. Churchill, Clifford G. Erickson, and Willard G. Warrington demonstrate that examinations can be effectively used for granting credit without any reduction in standards as well as producing a challenge for the able student to show his individual accomplishments.

The address by John E. Ivey, Jr., presented at the luncheon, challenged many of the traditional practices of higher education and suggested a number of steps which might be taken to improve

the quality of learning while taking care of more students and preserving, if not actually increasing, the time available to the faculty for research and scholarly endeavor.

The afternoon session explored the relationship between testing and evaluation and the developing conceptions of accreditation and institutional research. The papers by Dewey B. Stuit and James I. Doi demonstrated that the opportunities for imaginative contributions in these areas by measurement and evaluation people are many and the need for their contributions is great.

The papers of all participants were carefully prepared, well-delivered, and enthusiastically received. Henry Chauncey, President, and his associates at the Educational Testing Service are to be congratulated for having made possible one more conference of high quality and broad appeal. I am grateful to them all for having had the privilege of planning the sessions and acting as chairman.

Paul L. Dressel
Chairman

CONTENTS

3 Foreword by HENRY CHAUNCEY

5 Preface by PAUL L. DRESSEL

Session I

11 Factors of Verbal Achievement, JOHN B. CARROLL, Graduate School of Education, Harvard University

19 Schools of Thought in Judging Excellence of English Themes, JOHN W. FRENCH, Research Division, Educational Testing Service

29 Factor Analysis of Relevance Judgments: An Approach to Content Validity, LEDYARD R TUCKER, Department of Psychology, University of Illinois

Session II

41 Credit and the Advanced Placement Program, JACK N. ARBOLINO, Advanced Placement Program, College Entrance Examination Board

46 The Examining Program at Antioch College, RUTH D. CHURCHILL, Office of Testing and Evaluation, Antioch College

51 Examinations and Instruction by Television, CLIFFORD G. ERICKSON, Chicago City Junior College

57 Credit and Waiver Examinations at Michigan State University, WILLARD G. WARRINGTON, Office of Evaluation Services, Michigan State University

CONTENTS Continued

Luncheon Address

64 Management of the Learning Environment, JOHN E. IVEY, JR., President, Midwest Program on Airborne Television Instruction

72 **In Memoriam: Irving D. Lorge**

ROBERT L. THORNDIKE, Teachers College, Columbia University

Session III

75 Measuring the Quality of a College or University, DEWEY B. STUIT, College of Liberal Arts, State University of Iowa

87 The Nature and Use of Institutional Research, JAMES I. DOI, Institutional Research, University of Colorado

98 **List of Conference Participants**

Session

Theme:
Implications of
Factor Analysis for
Achievement Testing

JOHN B. CARROLL,
Graduate School of Education, Harvard University

Factors of Verbal Achievement

FACTOR ANALYSIS is indeed a perennial favorite as a subject for discussion at these conferences. I was asked to consider the implications of factor analysis for achievement testing, with special reference to skills and knowledges in the verbal domain. My title, "Factors of Verbal Achievement," may sound as if I am going to rattle off a catalogue of such factors. This is not what I intend to do. Instead, I want to consider some general questions about the role of factor analysis in studying and measuring achievement. I shall illustrate my points in connection with the verbal domain, but they are equally applicable in other domains. I shall, of course, restrict my attention to factor analytic work conducted by R-technique as applied to correlations of test variables over persons.

What does factor analysis really do? The pat answer is that it isolates a series of underlying "dimensions" of individual differences and makes it possible to describe a large set of test scores in terms of a relatively small number of such dimensions. Given a set of tests, the use of factor analysis helps one understand what kinds of things are being measured and provides an account of the "structure" of the domain under study. An attempt is made to "interpret" the separate factors or dimensions, and some of the more venturesome of the factor-analytic guild have even attempted to draw inferences from factor analytic results about the structure of the intellect itself.

I must say that I have been steadily losing sympathy for such a way of thinking about factor analytic results. The notion of "dimensions" may be a convenient one when we are trying to understand the geometric representation of the factor analytic model, but it makes less and less sense as a faithful way of describing behavior. It deflects attention from a better and more realistic description, a description, by the way, which may help in the

resolution of certain mathematical aspects of factor analysis.*

Essentially, any factor analytic study has to do with a series of behaviors or responses as embodied in a set of tests or other measuring instruments or procedures. The factor analytic procedure examines, as it were, all the possible pairs of these behaviors and identifies clusters of responses whose frequencies of joint occurrence in a given sample of individuals are greater than the frequencies of joint occurrence for other pairs of responses. The important datum is the frequency of joint occurrence of responses. The factor analysis might find, for example, that correct responses to two mathematics test items are more likely to co-occur than the responses to two items, one of which concerns mathematics and one of which concerns history.

There are a number of technical difficulties which I will pass over as not really relevant to my argument. It must be pointed out, for example, that most factor analytical studies do not study single responses, but rather, groups of responses embodied in tests. The net effect of this is artificially to construct metrics relating to the co-occurrence of certain sets of responses, the only justification being a certain significant degree of internal consistency among the responses in a set, as established by item-analysis techniques.

Factor analysis would be of more generalized import if it restricted itself to studies of sets of items considered singly, but this has usually been regarded as impractical because of the large numbers of items one would have to consider, and for other reasons. Another technical problem concerns the statistical treatment of joint occurrence data. One does not ordinarily enter joint occurrence data into a factor analysis, but instead uses some statistical measure of "relationship"; I have considered this problem in my recent presidential address to the Psychometric Society.[1]

I believe these technical problems, in any case, can be resolved in a way that will embrace the present analysis, in which I wish to emphasize that the essential data of factor analysis are the joint frequencies of pairs of behaviors, or classes of behavior, in a specified sample of individuals, and that the "factors" of factor anal-

---

* For example, the description urged in this paper would tend to minimize the importance of attempting to find any definite minimum rank for a matrix of correlations; rather, the question is one of ascertaining how many of the n dimensions into which a matrix can be factored are significantly associated with identifiable sources of variance. This description also supports the notion of correlated, nonorthogonal primary factors corresponding to correlated sources of variance which give rise to clustering within sets of responses.

ysis really correspond to sets of responses whose joint occurrences are conspicuous in comparison to those of other possible sets of responses. The "factor loadings" of a test variable on a factor may be regarded as being functions of the magnitudes of its joint frequencies of occurrence with other variables in the set.

When we try to "interpret" a factor, therefore, we should attempt to consider the source of the frequencies of co-occurrence. That is, why do the responses clustering in a factor appear together?

There are many possibilities. Among these are: 1) the learning or the performance of one response is prerequisite to, or implied in, the performance of another (this is most frequently the case where it is found that the items can be scaled in the Guttman sense); 2) the learned behavior represented in one response has transferred to the other response; and 3) because of the accidents of personal history or because of the common experience of certain numbers of the group of persons, there was a higher probability that both of any pair of responses were learned together than that either would have been learned alone.

Note, now, that this last possibility makes no appeal to any psychological considerations about the two responses; it merely refers to the actuarial fact of their co-occurrence, beyond chance, in the particular sample under study. I suspect that this account will often apply to the sets of responses included in many so-called "achievement" tests. And if it is correct, a demonstration of it could be brought about on an experimental basis. Indeed, given a series of behaviors, one could arrange a training situation so as to obtain any particular factor structure that one might desire. For example, suppose I have 50 randomly assorted nonsense syllables, and suppose that I wish to establish a training program so that there will be five factors of achievement, each represented by a specified set of 10 nonsense syllables. All I have to do is to arrange my 50 syllables into the five lists of 10 nonsense syllables apiece and then proceed to give the subjects in my experimental group different amounts of training on the 10 lists. For example, random halves of the group would get either much or little training on the first list, different random halves would get much or little training on the second list, etc. This would be an easy experiment to do; it might truly be worth doing if it would enlighten anybody.[2]

The model developed here is similar, I think, to any situation in which one analyzes a series of achievement tests. An appropriately designed factor analysis study in a high school could yield an English factor, a Social Studies factor, a Physics factor, a

Chemistry factor, and perhaps even a Driver Education factor, among others. These factors could exist solely by virtue of the fact that these courses of study consist of a series of responses which are learned together (or not learned together, in the case of the failing students).

Actually, there are few factorial studies which have been restricted to "achievement tests" covering distinct courses or syllabi, and therefore it is difficult to cite cases illustrating the design described here. No matter—the same type of interpretation in terms of joint learning of responses can be applied to factorial studies of aptitude tests, intelligence tests, and the like, for all of these can be regarded as achievement tests of a special kind. It must be remembered that the distinction between achievement tests and other types of tests is largely one of purpose, intended use, and scope of content rather than any qualitative difference in type of content. If, therefore, I now turn my attention to a scrutiny of some of the better established "factors" of intelligence and aptitude tests in the verbal domain, it will now be understood that my title, "Factors of Verbal Achievement," is not inapplicable. Some of the mystique which has surrounded factor analytic results needs to be torn away. "Factors" as entities have been overhypostatized. They cannot and must not be regarded as putative "things" which somehow exist in whatever bodily structures may be responsible for behavior.

Take, for example, the question which has sometimes been formulated as that of whether abilities tend increasingly to specialize with age. Some evidence has been interpreted as showing that they do. If so, is it not better to consider this result a manifestation of the fact that in our culture the learning experiences of individuals tend to become gradually more and more heterogeneous?

Another example is afforded by the apparent contrast between British and American results in the study of intelligence tests. If British results have found a larger place for a general intelligence factor *g* than American results, could this not be a reflection of a greater uniformity of learning experiences in British grammar schools—uniformity, that is, in content, if not also in teaching method or quality of learning?

I offer these only as speculations; there exists no clear-cut evidence for either of them.

Let us turn our attention, then, to some of the factors of verbal achievement. The basic hypothesis we shall be examining is that whenever we identify a "factor" through factor analytic procedures, the source of this variance is to be found in differential learning

experiences. Actually, in this brief talk I can do little more than to suggest how this hypothesis could be investigated.

Take first the well-known *V* factor, the "verbal knowledge" factor. A variety of tests—all essentially "achievement tests"—seek to determine whether various verbal responses have been previously learned. I have in mind vocabulary tests, verbal comprehension tests, grammatical usage tests, and the like; morphological, syntactical, and lexical aspects of language are all involved, as well as certain kinds of language habits which are better studied in terms of conditional probabilities than in terms of linguistic analysis.

There are many obvious reasons for the joint learning or nonlearning of the responses involved in these tests. One stems from the fact that the members of the English-speaking community are exposed to widely varying amounts and qualities of English. Among the influences which determine the amount and quality of English to which a child or adolescent will be exposed are the following: socio-economic class, language ability, and occupational status of parents; rank among siblings; sex; and type of child rearing. In addition, there are the person's (learner's) interests and temperament which engender more or less exposure to and handling of verbal materials.

In any case, the very structure of language insures a clustering of the responses in a factorial sense. In a representative group, the joint frequency with which any two verbal responses will be present will be greater than the joint frequency of any one verbal response and any nonverbal response (for example, a mathematical or a spatial response).

It is strange, but true, that even now we can cite no thoroughgoing, all-inclusive factor analytic study of all the ordinary English language skills; therefore we have had to piece together our knowledge from a variety of studies. It is now pretty certain, however, that there are separate "factors" for the traditional language modes, that is, reading, writing, speaking, and listening.

Reading, whether measured in terms of speed or of comprehension, is probably closest to the traditional *V* factor; that is, it is highly correlated with the *V* factor and in many studies indistinguishable because of lack of clear study design. It is reasonable for reading ability to be factorially similar to verbal knowledge because verbal knowledge tests tap the kinds of responses which are more likely to be learned through reading than through listening. Furthermore, most verbal knowledge tests are printed tests that require reading.

Listening ability, studied extensively in a recent thesis at Harvard by Spearritt, seems to be fairly well distinct from verbal knowledge and reading ability, even though correlated.[3] Apparently the learning experiences that make one able to listen to auditorily presented verbal material are sufficiently specific as not to correlate highly with reading comprehension. It is quite plausible that individuals have different degrees of experience in listening to and understanding material presented in oral form. And if you are having difficulty following me right now, this is perhaps a reflection of the fact that your learning experiences are different from mine!

Recent studies at Harvard have also shown that there are several factors in the domain of speaking ability, keyed to the various types of situation in which speaking is required—ranging from private conversation and interviews to public speaking occasions.[4] Learning experiences in these several situations vary, it would seem, giving rise to distinct factors.

Much the same sort of situation may exist in the field of writing ability, which has perennially been regarded as one of the most elusive things to measure. That is, writing ability may be a matter of situational control, with achievement varying according to the kind of task which is set.

Insofar as good writing depends upon knowledge of language responses—both morphological, syntactical, and lexical—it will be found related to the verbal knowledge factor V; indeed, in some College Board studies it appears that verbal knowledge tests are more satisfactory measures of writing ability than writing tasks themselves. We have not yet been able to measure the reasoning and organizational aspects of good expository and creative writing independently of writing itself, perhaps because the two are inextricably bound together. The many studies showing that the teaching of formal grammar is practically irrelevant to writing ability suggest that formal grammar knowledge constitutes a set of achievements attained quite apart from experience in writing.[5] Strangely enough, I do not know of any factorial studies which have concerned themselves with formal grammar, as distinct from the usual type of grammatical usage test. There is much investigation yet to be done in the realm of grammar and its relevance to English composition, the learning of foreign language, etc. I would suggest further experimentation with the kind of "grammatical sensitivity" test which is to be found in my Modern Language Aptitude Test, because it seems to me that this test, which measures ability to identify grammatical functions and patterns

(without any knowledge of formal grammar or terminology) may be of greater relevance to the kinds of responses which are called for in English composition.[6]

The so-called Word Fluency Factor *W* is an interesting example of a factor which has been masquerading as some sort of special ability but which, I believe, can in large part be interpreted as the result of certain kinds of learning experience. Tests of the *W* factor, you will recall, are the ones that require the subject to rearrange letters in anagrams tests, to think of words beginning and ending with certain combinations of letters, and the like. Is it not possible that an individual's experiences with the phonetic correspondences of letter and letter combinations—that is, with "phonics"—may have much to do with his ability to perform the kinds of tricks demanded on tests of the *W* factor? There exist tests of "phonic skill" which, one may hypothesize, would show a substantial correlation with tests of the *W* factor.[7] The problem is to identify the sources of the learning experiences which are responsible for performance on these tests; perhaps they could be identified only by finding out how the individual learned to read. Was he taught phonic skills, or was he allowed to learn to read primarily by whole-word recognition techniques?

In some fields of verbal achievement, it would appear that learning experiences are rather highly specific to certain kinds of situations. Take, for example, the "Naming" factor which has been identified by me and several investigators.[8] It appears whenever the subject is required to give names to a series of stimuli. Can it be that such an ability—the ability to do this rapidly, that is—is acquired as the result of specific practice, perhaps in early childhood? It would take rather an elaborate longitudinal study to find out, I am afraid. Similar remarks could probably be made for a number of other factors of verbal ability—for example, Expressional Fluency, Associational Fluency, and Ideational Fluency as described, for example, by John French in his kit of reference tests for factor studies.[9]

Of course, even after we have accounted for as much of the variance in all these factors of verbal achievement as possible on the basis of differential learning experiences, there might be some residue traceable to genetic or constitutional factors. It would be satisfying, however, to get these factors solidly tied in with the processes of learning. Even more satisfying would be to find out to what extent these factors are subject to training, that is, to post-test learning experiences. This is an exciting prospect for future experimental work. At the moment, one cannot find much

evidence of activity in this area. It is perhaps regrettable that we are all just too busy making and analyzing tests to leave time for that.

REFERENCES

1. Carroll, John B., "The Nature of the Data, or How to Choose a Correlation Coefficient." *Psychometrika*, 26, December, 1961.
2. Skager, Rodney W., "The Effects of Two Training Procedures on the Differentiation of Numerical Ability." Unpublished Ph.D. dissertation, University of California at Los Angeles, 1961.
3. Spearritt, Donald, "A Factorial Analysis of Listening Comprehension." Unpublished Ed.D. dissertation, Harvard University, 1961.
4. Marge, Michael, "Home Background Influences on the Development of Oral Communication Skills in Children." Unpublished Ed.D. dissertation, Harvard University, 1959.
5. Strom, Ingrid M., "Research in Grammar and Usage and Its Implications for Teaching Writing." *Bulletin of the School of Education, Indiana University*. Vol. 36, No. 5, September, 1960.
6. Carroll, John B. and Sapon, Stanley M., *The Modern Language Aptitude Test*. New York: The Psychological Corporation.
7. Brown, Grace M. and Cottrell, Alice B., *Stanford Diagnostic Phonics Survey*. Palo Alto, California: Consulting Psychologists Press.
8. Carroll, John B., "A Factor Analysis of Verbal Abilities." *Psychometrika*, 6, 279-307, 1941.
9. French, John W., *Kit of Selected Tests for Reference Aptitude and Achievement Factors*. Educational Testing Service, 1954.

Schools of Thought in Judging Excellence of English Themes

JOHN W. FRENCH,
Research Division, Educational Testing Service

A PERENNIAL bone of contention between English teachers and psychometricians has been the effectiveness of essay tests of writing ability. The testing program of the College Entrance Examination Board has reflected the crazy ups and downs of this controversy. Fifteen years ago a written composition was an integral part of the Board's achievement test in English. The psychometricians then proved that a verbal aptitude test could predict English grades or ratings on writing ability better than a test which actually required the students to write.

The College Board then dropped the English composition from its program. Shouts of horror arose, and a committee was formed. This resulted in a new, ingeniously garnished two-hour essay test called the General Composition Test. The College Board, with some misgivings, put that into its program. The psychometricians, of course, swung into action and came up with lots of data to show that even this new test when compared with the verbal aptitude test had low reliability, low validity, high cost, and no important effect, during at least a five-month period, on either what the English teachers taught or what the students studied.

The College Board then dropped the General Composition Test from its program. Shouts of horror arose, and a committee was formed. At the present time, the College Board again has an English Composition Test, called the "Writing Sample" in its program, but this one doesn't get scored at all. We just send carbon copies to the colleges, and wonder what the colleges are going to do with them and how much longer they are going to want to receive all those almost unreadable carbon copies.

Well, enough about the controversy. Let's consider the sources of error in the scores on a composition test. It is possible to divide the error into four parts:

1. Student Error. A student can do well one day and can do

poorly another day on the very same task. This is, of course, true of both composition and objective tests, but can be greatly aggravated for the composition test, because the student may guess sometimes correctly and sometimes incorrectly about what style of writing or what kind of content his particular examiner will want to find.

2. Test Error. A test calls for a sample of the student's behavior. A long objective test calls for many small samples or items; the items that happen to favor one group of students are usually balanced by items that are especially difficult for the same group. The composition test is almost like a one-item test. Some students may happen to enjoy the topic assigned, while others may find it difficult and unstimulating; this results in error.

3. Scale Error. The reader of a composition can be an easy marker or a tough marker. Therein hangs much of a student's fate. To get all readers for an examination to grade papers on the same scale is no easy matter. It may be largely a matter of administrative persuasion.

4. Reader Disagreement. Even if readers could be persuaded to use the same scale, that is, to give the same proportions of A's, B's, C's, etc., their grades would still not look alike, because there is disagreement on what kind of writing is best.

To recapitulate, an essay test is subject to student error, test error, scale error, and reader disagreement. The experiment I am about to discuss is an exploration of this fourth kind of error: Reader Disagreement, or, in more elegant terms, the dimensionality in the grading behavior of the readers. The study was carried out by Paul B. Diederich, Sydell T. Carlton, and the speaker with the support of the Carnegie Corporation.

The stimuli for the readers were 300 short essays written as homework by college freshmen. These were written on two rather general essay topics: "Who Should Go to College?" and, "When Should Teen-agers Be Treated as Adults?" The papers were selected from a larger number of available papers so as to include those written by a larger percentage of low ability and very high ability students than is usually found in a freshman English course. This was done to make discrimination among the papers easier, so that the different kinds of discrimination made by the readers would stand out as clearly as possible.

We considered our interest in the evaluation of writing to extend beyond teachers of English. We had in mind all kinds of persons who have reason to savor or suffer from the quality of writing achieved by graduates of our colleges. Consequently, we

sought and were able to secure the participation of ten English teachers, nine social scientists, eight natural scientists, ten writers or editors, nine lawyers, and seven business executives—all highly competent people in their respective fields. All of these—bless them—read and graded 300 student essays for what can only be considered nominal compensation.

The instructions for the grading that were given to the readers were probably quite important to this study. In the past, when the College Board had readers grade an essay test, every effort was made to bring about high agreement on the grades that were assigned. The readers would be assembled; they would discuss and agree to certain rules for grading students' responses to the particular essay topic; they would grade sample papers, discuss the results, and revise the rules to cover all contingencies. After three days of this they would start doing the grading that counted. The average correlation of one reader's grades with those of another would be about .70. That would not be bad, if we could be sure that the things the readers were able to agree on were just as important as the things they must ignore in order to achieve this high level of agreement. In the present study, the readers were not assembled to discuss their ideas about reading. They worked at home. They were told merely to "Use your own judgment as to what constitutes 'writing ability.' Do not assume that we want you to do this or that." Nothing was said about style or grammar or quality of ideas. They were merely asked to "sort the papers into nine piles in order of merit," with at least six out of the 150 papers on each topic in each pile. In addition, they were asked to write on the papers comments as to why they liked or disliked them.

The next step in our procedure was to compute the correlations between every possible pair of readers. In spite of the fact that discrimination was aided by the wider-than-normal range of ability among the students, the average correlation was only .31. You can understand just how low this is, when I tell you that, out of the 300 papers, 101 received all nine different grades, and no paper received less than five different grades.

It was the objective of this research to analyze the agreement, such as it is, and to find out whether schools of thought exist among the readers. One hypothesis was easy to check by looking at the correlations. Do English teachers represent one school of thought, social scientists another, and so on? If such is the case, the average correlation among the English teachers would be higher than .31. Since the College Board readers are all English teachers, this might account for part of the success in achieving

an average correlation of .70 in the Board program. We then computed the average of our correlations between English teachers and found it to be .41. Since this is appreciably above the average for the whole group, there seemed at first to be some evidence for an English-teacher school of thought. However, we computed similar averages for the other occupational groups, and these were all very low. It turns out that members of all the other occupational groups agreed better on the average with the English teachers than they agreed with members of their own group. This means that the English teachers do not represent a school of thought but are able to grade papers with less random error than that which prevailed in the other groups.

In order to find out what schools of thought do exist among readers, the matrix of intercorrelations among the readers was submitted to a factor analysis. Since it is possible to consider the grades of each reader as an evaluation of the student writing the essay, we thought it reasonable to include in the factor analysis other available evaluations of the students: their College Board Scholastic Aptitude Test (SAT) Verbal scores, SAT-Mathematical scores, and English Composition Test scores. These scores were included just as if there had been three extra readers.

Six principal-component factors were extracted, the first being very large in comparison to the others. This indicates that there are six sources of variation or schools of thought explaining the relationships among the readers and the tests. However, the large size of the first factor indicates that such agreement among readers as exists is represented by a set of values having importance for all readers. This set of values could have been defined by a general factor, which would be supplemented by the five other group factors, representing differences of opinion beyond the central core of agreement. However, in order to obtain the clearest possible picture of the different opinions of the readers, the six factors were rotated to oblique simple structure, the position of each rotation being determined graphically.

It became apparent immediately that one factor was a test factor. The three objective College Board tests appeared close together and quite remote in the six-dimensional factorial space from any of the readers. There was nothing in this particular factorial space to separate the mathematical test, SAT-M, from the verbal tests, SAT-V, and the English Composition Test.

Each of the other five factors was represented by a small group of readers having high loadings on it. Here commenced the most challenging phase of this research. Since, in this study, the factors

are groups of readers, it might be possible to find something in their backgrounds or experiences that was responsible for the groupings. This would lead to a very different kind of analysis from the one we carried out. Our concern was directly with the behavior in grading papers that caused the factors to emerge. Nevertheless, it was sensible first to check two obvious characteristics of the readers: occupation and sex.

Table 1 gives the occupational fields of the readers having factor loadings of .25 and above. The mixture of occupations on every factor leads us to reject the hypothesis that schools of thought in grading papers are mediated by occupational affiliations. It can be considered noteworthy, however, that the three highest loadings on Factor III are all from the one group of writers and editors, and more than half of the readers listed for Factor IV are English teachers.

Four of the 53 readers were women. Three of them had their highest loadings on Factor V, but none of these loadings were high enough to seem convincing; they were .30, .22, and .20. The fourth woman reader did have a high loading, .40, but it was on Factor I. Thus, we were able to identify no sex factor.

Table 1

The Fields of Readers Having Factor Loadings of .25 and Above

(Three highest loadings underlined)

FACTOR I		FACTOR IV	
English	.33, .33, .30, .27	English	.47, .36, .35, .32, .31, .30, .30
Soc. Sci.	.58, .53, .25	Soc. Sci.	.26
Nat. Sci.	.47, .34	Nat. Sci.	.38
Writers	.40, .37, .31, .27	Writers	.29
Lawyers	.41	Lawyers	.33, .26
Business	.44, .43	Business	.40
FACTOR II		**FACTOR V**	
Nat. Sci.	.42, .31	Soc. Sci.	.35, .29
Writers	.35, .35	Nat. Sci.	.29, .25
Lawyers	.37	Writers	.26
Business	.37, .32, .29, .29	Lawyers	.30, .25
		Business	.31, .27
FACTOR III		**FACTOR VI**	
Soc. Sci.	.26	SAT-V	.67
Nat. Sci.	.27	ECT	.65
Writers	.41, .34, .33, .28	SAT-M	.58
Lawyers	.26		

Another approach to the interpretation of the factors was to observe various objective characteristics of the papers and correlate these with the grades assigned to the papers by readers representing each factor. This was done by Dr. Paul Jacobs. Among the objective characteristics tried were length in inches, proportion of unusual words as defined by the Lorge magazine list, and number of mechanical errors—not errors on which readers happened to comment but a complete listing of all identifiable errors compiled by independent specialists. The only one of these that worked was *length*. Average grades of the five readers with highest loadings on Factor I correlated about .8 with length. Average grades of readers with the lowest loadings on Factor I correlated about —.3 with length. This fitted well with some preconceptions many of us had about some readers of English essays, but we were later to change this interpretation, and possibly, this preconception.

A third and more thorough approach to the interpretation of factors was through the comments that were written on the papers by the readers. This is a study with a lot of data. There were 53 readers, 300 papers each, and an average of about 11 words commenting on each paper: some 50,000 comments totalling 175,000 words. Therefore, we limited the analysis to papers bearing the three highest or three lowest grades by the readers who had the three highest or three lowest loadings on each factor. This called for the analysis of 11,018 comments written on 3,557 papers.

After a prolonged study of the comments without reference to the factor loadings of the readers, 55 categories of comments were adopted and listed on scoring sheets for a systematic tallying of the kind of comment made by each reader in the analysis group. All classifying and tallying of comments was done by a single person who knew nothing of the factor loadings of the readers. For each comment, the length of the comment was recorded, and a tally mark was entered for all categories covered by the comment. For example, if a reader said of a paper, "Good ideas and organization but several spelling errors," tallies were entered opposite "Quality of Ideas," "Organization," and "Spelling." If the reader went on to encircle two spelling errors, he was given two additional tallies opposite "Spelling."

It was now possible to compute the percentage of a reader's comments that were devoted to comments in each of the 55 categories. This percentage was computed separately for each of the two essay topics, and the readers were found to be consistent in the nature of their comments from topic to topic. Now, by sum-

ming the percentages for each topic for each of the three readers who were highest on a factor, we arrived at a figure we called the *index of emphasis* for the factor. For comparison, an *index of emphasis* was also computed for the three readers who were lowest on each factor.

These *indices of emphases* for 33 of the 55 categories are shown in Table 2. Each category is listed with the factor for which the three readers with highest loadings had the greatest *index of emphasis*. For example, in the first row of the table, Relevance is listed for Factor I, because the three readers with highest loadings on Factor I used this category a higher percentage of times than did the readers who were highest on any other factor. The lowest readers on each factor did not affect the listing. Within each factor the categories are listed in order of diminishing sums of the indices over all five factors. Twenty-two categories have been omitted from the list entirely because their indices were so low that their placement would not be reliable.

The factors, or schools of thought among the readers, may now be interpreted by considering the categories listed for each factor and by referring judiciously to the relative sizes of the various indices. Our interpretations are given in capital letters as names for each factor. They are, IDEAS, FORM, FLAVOR, MECHANICS, and WORDING. It is now of interest to recall the close association, mentioned earlier, between grades of readers representing Factor I and the length of the essays in inches. A reasonable explanation seems to be that ideas, particularly quantity of ideas, requires room on the page.

Now, in this analysis, in order to achieve the clearest, truest picture of the factors, we rotated the factor axes obliquely. This means that the factors are not independent from each other. The intercorrelations among them are given in Table 3. You will see that the MECHANICS and WORDING factors are rather closely associated with each other and also have some high correlations with the other factors. Perhaps the most interesting finding in Table 3 is the row of correlations for the test factor. The test factor is unrelated to IDEAS and essentially unrelated to FORM and FLAVOR. On the other hand, there is a substantial relationship to MECHANICS and WORDING. This tells us something about the tests, something that all of us may not be happy to learn. The test scores are somewhat like the essay grades given by readers of the MECHANICS and WORDING schools of thought, and they appear to be completely unrelated to the grades assigned by readers who grade on ideas.

Table 2

Index of Emphasis for the Highest and Lowest Readers on Each Factor

Categories of Comments	Three highest readers I	II	III	IV	V	Three lowest readers I	II	III	IV	V
I. IDEAS	*									
Relevance	44	26	29	15	34	16	40	23	20	29
Concise-wordy	20	15	03	17	18	22	13	19	24	09
Clarity of ideas	18	10	16	09	15	07	35	11	06	16
Quantity of ideas	28	09	01	20	08	09	30	13	08	14
Development	19	18	04	12	09	13	11	12	03	05
Too brief or long	21	13	02	10	14	06	07	01	06	02
Persuasiveness	16	00	07	03	06	06	20	03	06	06
Ending	13	04	03	03	09	07	10	06	03	09
Generality	09	08	00	03	03	00	03	11	00	05
II. FORM		*								
Spelling	58	105	06	73	72	96	11	48	07	31
Clarity of expression	39	65	57	62	25	85	58	17	12	33
Organization	24	44	35	09	26	15	19	52	23	15
Coherence of ideas	21	34	19	17	29	22	32	31	21	17
Reader agreement	10	39	05	23	08	24	25	07	16	10
Analysis	13	35	03	04	17	01	08	42	19	10
Maturity	07	09	01	08	07	06	08	12	05	06
III. FLAVOR			*							
Quality of ideas	72	51	89	30	43	51	64	51	26	25
Style (general)	19	05	77	08	19	32	14	13	31	09
Mechanics (general)	19	01	57	19	00	01	26	14	01	07
Originality	15	08	39	13	16	07	24	33	05	04
Interest	08	02	32	10	15	04	11	06	04	02
Beginning	03	03	20	01	12	01	01	05	07	01
Sincerity	05	06	12	00	09	04	06	04	02	04
Info. & Illus.	06	01	08	01	06	05	01	10	05	00
IV. MECHANICS				*						
Punctuation	10	02	00	69	29	17	06	23	15	17
Grammar	17	17	02	39	22	34	08	14	17	23
Sentence structure	01	04	00	21	21	11	06	03	14	02
Phrasing, idiom	10	03	00	14	05	18	04	10	08	18
V. WORDING					*					
General	20	35	45	20	49	10	52	49	33	04
Word choices	18	07	04	32	34	43	13	26	88	44
Logic	10	10	09	05	11	03	10	14	01	06
Clichés	06	03	10	04	10	06	11	01	02	08
Jargon-slang	03	02	00	03	03	04	03	00	00	05

* The column in which the highest readers on each factor stood first is marked by an asterisk.

Table 3

Factor Intercorrelations

	I	II	III	IV	V	VI
I IDEAS		.03	.27	.25	.55	−.02
II FORM	.03		−.03	.33	.10	.16
III FLAVOR	.27	−.03		.45	.34	.07
IV MECHANICS	.25	.33	.45		.63	.50
V WORDING	.55	.10	.34	.63		.45
VI TESTS	−.02	.16	.07	.50	.45	

Isn't this what the English teachers have been telling us psychometricians all along—that essay tests of writing ability measure something that the objective tests do not measure? They seem to be right. Form, flavor, and, particularly, ideas are missed by the tests. But, then, of course, they are also pretty much missed by the readers. The low correlations among the readers and the 101 papers receiving all nine grades can only mean that nothing at all is being measured very well by the essays. It could be pointed out that readers trained to grade a particular essay can achieve a figure of .70 for correlations among themselves. But are they still grading on ideas, form, flavor, or whatever they want? Of course, they are not. The training forces them to grade in ways which they would not ordinarily grade. In fact, the *well-trained* reader may simply be measuring verbal aptitude—poorly.

What, then, can be done about these interesting things that essays can rather foggily measure? A first step will be to decide what we *want* to measure. Let's say it is ideas, form, and flavor, *not* mechanics, wording, or scholastic aptitude. The present study points to certain individuals who chose without special direction to pay particular attention to ideas, form, and flavor. Right here there is an advance in the reliability of measurement. The average intercorrelation among the seven readers whose loading was above .40 on IDEAS, for example, was .46, as compared with .31 reported earlier for the whole group of 53 readers. Now let's increase agreement still farther by finding a group of readers who at least espouse the importance of reading for ideas, form, and flavor. Let's have these readers choose the essay topic for sensitivity to these particular qualities, come to an agreement on how to grade the chosen topic, and train themselves to grade it in a consistent way. There can still be no guarantee that agreement will go as

high as .70, because the qualities being graded may be harder to grade consistently than something like mechanics. However, the readers, at least, would be allowed to grade in more nearly the way they like best and that should help reliability.

In addition, a clear perspective of what is being looked for by the readers might have a diminishing effect on student error. With some teachers stressing one quality and others harping on another, a student may very reasonably respond to a test essay by stressing his own teacher's favorite essay characteristic, or he may allow the essay topic to suggest one thing or another to stress. Only if he is lucky will he even try to impart to his essay the characteristic that his particular reader is going to inspect most closely. On the other hand, if the readers were determined to grade on, say, an equal weighting of ideas, form, and flavor, and if they communicated this intention to the students through test directions, it is likely that the students would respond by a more uniform effort to stress these things in their writing. The test would now measure the students on something they have done more nearly to the limit of their capacity, and so would benefit by reduced student error.

So, if we psychometricians can encourage testing and further clarification of those aspects of writing that objective tests cannot measure, encourage the use of readers who favor grading those particular qualities that are desirable to grade, and see to it that the students are aware of what they are being graded on, we can enlighten rather than merely disparage the polemic art of essay testing.

Factor Analysis of Relevance Judgments: An Approach to Content Validity*

LEDYARD R TUCKER,
Department of Psychology, University of Illinois

CONTENT VALIDITY seems to be an almost universally accepted requisite of tests used to evaluate achievement and performance. For example, *Technical Recommendations for Psychological Tests and Diagnostic Techniques* states: "Content validity is especially important in the case of achievement and proficiency measures." Similar sentiment is expressed in *Technical Recommendations for Achievement Tests*. Texts and manuals on the construction of achievement tests emphasize the importance of consulting subject-matter experts so as to achieve not only accuracy in the questions included in a test but also a proper coverage of material in the area to be examined.

Problems exist, however, in the precise definition of content validity. A frequent definition is in terms of the sampling of a specified universe of content. While a coefficient of content validity could be developed theoretically, relating performance on a test to performance over the universe of content, the Achilles Heel in this definition is the specification of the universe of content. Some universes of content are easy to specify. A universe for spelling proficiency, for example, might contain all the words in some accepted dictionary. Other universes, such as the composition of business letters, are less easily specified. A unit task in this case may be taken as the composition of a letter in a given business situation and the universe of content can be defined in terms of the universe of business situations. But this universe may be infinitely varied with a large number of common situations and a variety of increasingly rare situations. Complete enumeration of

* This paper is based, in part, on research conducted by Dr. William E. Coffman of Educational Testing Service and the author which was jointly supported by Educational Testing Service and the College Entrance Examination Board.

the spelling proficiency universe is quite possible while for the business letter such an enumeration is quite impossible.

Fortunately in the case of the business letter, a universe of business situations does exist in the real world and a survey of these situations is possible. Such a survey could yield a sample of situations to be used in a test.

There exist in education, however, many cases for which the specifications of universes of content are much more difficult. Often a general field of knowledge or skill is defined in general terms which correspond to no existing universe of situations. These terms of definition may be such that qualified people could judge whether a given situation lies within the universe or does not. Frequently, however, the general terms of definition are open to different interpretation by qualified people. This leads to proper controversy as to the nature of the universe of content for each such field of knowledge or skill and compounds the problem. There exists neither a universe in actual existence nor agreement whether particular situations lie in the universe or not. Consider, for example, a general science course for ninth graders. Suppose that this course is described as presenting an introduction to scientific method and knowledge in the fields of physical and biological sciences. Insofar as the totality of existing scientific knowledge and procedures is growing, there is considerable question whether a universe of these situations is in actual existence. To guarantee complete existence of a universe of content, the general terms should be qualified with a statement as to present knowledge. Even with this statement, the universe for a general science course is not completely delineated. What material is introductory and what is not? Answers to this question depend on the judgments of the person answering the question.

A commonly stated proposition is that the objectives of instruction should be described in concrete terms rather than in vague abstract terms. In many aspects this description in concrete terms is equivalent to enumeration of the universe of content. Once a complete description of the objectives of a course is made in concrete terms a considerable part of the problem of content validity of a measuring instrument for evaluation of the achievement of students has been solved. A close inspection of this procedure of translating the abstract terms of a concept into concrete terms reveals again the possibility of differences of opinion among individuals, even among individuals commonly accepted as qualified in some particular field. Strong controversy exists, for example, as to whether in citizenship training students should be trained

toward political conservatism or toward political progressivism. There would not be agreement between individuals advocating one or the other of these views as to the concrete content of a course in citizenship although they might agree that such a course under the general description would be worthwhile.

From the foregoing, we may conclude that the content validity of a test not only depends on the material included in the test and a general definition of an area of knowledge and skills when this definition is couched in abstract terms, but also depends on the point of view of the person interpreting the general definition of the area. A test having content validity for one interpretation may lack content validity for another interpretation. Herein lies a major problem in the construction of standardized achievement tests. When such divergent interpretations occur, no single test yielding a single score can have content validity for all interpretations and several tests are indicated, one for each interpretation. Thus, it is of vital importance that systematic knowledge exist as to the extent and kind of different interpretations of accepted general concepts of areas of knowledge.

In some cases the divergent interpretations are well known, especially when each view has been vigorously advocated by some individuals. In other cases the divergent interpretations may not be well known, especially when there has not been active debate. The divergent interpretations may not be solidified into distinct points of view, rather, there may be flexible discussions and disagreements couched in polite tolerance. Experimental methods for investigating these cases which yield indications of the latent controversies and divergent interpretations should add materially to our ability to construct tests with content validity. One such procedure is involved in the factor analysis of relevance judgments of items. I shall discuss this procedure as studied in research by Dr. William E. Coffman and the author which was supported by the College Entrance Examination Board and the Educational Testing Service.

As part of its continuing program of test research and development, the College Entrance Examination Board, a few years ago, appointed committees to work with the Test Development Division of Educational Testing Service in the preparation of experimental tests in three broad content areas: science, humanities, and social studies. These tests were termed "The Tests of Developed Abilities" and were to examine college applicants on the broad knowledge and skills that might be expected to grow out of experiences in a sound program of secondary education. Each committee, includ-

ing the Social Studies Committee, was responsible for defining "developed abilities" in its field and for constructing test questions to measure the abilities thus defined. The Social Studies Committee defined its field through an outline naming the general abilities which the student should have mastered and the content through which the abilities should be applied. The concrete definition consisted of 225 test questions. A majority of the questions were of a "composite" type, requiring the simultaneous use of both knowledge and skill, but two other types were included. One of these types, termed "information items," focused primarily on knowledge while the other, termed "self-contained items," required only that the student do something with content presented explicitly in the item.

But would a test score derived from the 225 items be valid as a measure of developed abilities in social studies as judged by a sample of experts more representative of social studies teachers than the five-member committee? To answer this question, the College Board appointed a panel of critics to review the proposed test material. Data for our study were collected from 17 members of this panel. We wished to investigate for systematic differences in value opinions regarding the proper content of such a test.

Each judge was requested to review each item in the sample of 225 and to rate it on a scale from 0 to 5 as to importance in evaluating the developed abilities of the examinees. The panel members were further warned to discount the effect of difficulty of an item in making the judgment as to importance. They were asked to judge difficulty separately and then to remember that while one might attach great importance to a right answer to a difficult item, a similar importance could not be attached to a wrong answer. Likewise, while one might attach great importance to a wrong answer to an easy item, a similar importance could not be attached to a right answer. And finally, both right and wrong answers might be considered important when items were of middle difficulty.

The task, then, was to evaluate the relevance of each item for measuring the developed abilities in social studies. A rating of 0 was defined to indicate that the item was irrelevant, and a rating of 5 was defined to indicate that the item was extremely relevant.

A sample of the ratings is given in Table 1. The order of the judges and of the items is related to results obtained in the study and will be discussed later as will the last two columns of this table. Note that each of the rating values from 0 to 5 is used, especially from 1 to 5 with an occasional 0. In fact, the ratings

Table 1

Ratings of Judges on a Sample of Items

Item No.	Judges																Group	
	5	14	3	4	7	6	9	1	16	2	8	10	12	13	15	17	11	A (5, 14, 3)	B (12, 13, 15, 17, 11)
3	2	1	2	3	2	1	5	2	3	3	1	3	4	3	3	3	4	1.7	3.4
46	1	3	3	5	2	3	4	2	4	3	4	4	4	4	3	3	5	2.3	3.8
10	1	1	2	4	2	3	2	2	3	2	1	3	4	2	2	2	3	1.3	2.6
5	2	3	1	3	2	2	3	2	3	4	1	3	4	2	2	2	4	2.0	2.8
2	3	2	2	3	2	1	5	4	3	3	2	3	4	3	3	3	4	2.3	3.4
26	2	4	4	5	3	4	3	4	4	4	3	4	4	4	5	3	5	3.3	4.2
4	3	3	2	3	2	1	3	3	2	2	1	3	4	3	2	3	3	2.7	3.0
36	2	3	2	5	2	4	3	3	4	4	4	2	3	2	4	2	4	2.3	3.0
40	3	3	4	3	2	3	3	4	3	4	4	3	3	2	4	3	5	3.3	3.4
43	2	3	4	3	2	2	2	4	4	4	4	3	3	3	4	3	5	3.0	3.6
39	4	4	4	4	2	5	3	2	3	3	4	4	5	3	5	3	5	4.0	4.2
16	3	2	3	4	2	2	3	3	2	5	1	2	3	2	1	0	5	2.7	2.2
32	3	4	4	5	2	4	2	4	4	4	3	3	5	3	5	1	4	3.7	3.6
50	3	4	4	3	3	2	4	4	4	4	4	4	4	0	4	4	5	3.7	3.4
51	3	4	5	4	1	3	4	4	4	4	3	1	4	3	0	4	5	4.0	3.2
81	4	3	5	4	3	2	4	2	4	3	3	3	3	3	2	1	5	4.0	2.8
65	2	3	4	4	2	1	3	3	3	3	1	3	2	3	2	1	1	3.0	1.8
80	4	3	5	3	3	3	4	3	3	3	4	3	3	2	1	2	5	4.0	2.6
52	3	5	5	4	3	4	4	5	3	3	5	1	4	0	4	3	5	4.3	3.2
85	4	4	5	2	3	0	4	3	2	3	2	3	4	3	1	1	2	4.3	2.2
110	5	5	5	5	3	2	3	4	4	3	4	3	4	1	2	3	5	5.0	3.0
104	4	4	5	5	2	1	3	3	2	3	3	3	3	3	1	0	3	4.3	2.0
95	4	4	5	5	2	4	3	2	3	3	3	2	3	1	2	0	1	4.3	1.4
199	5	5	5	5	3	4	3	3	4	3	2	3	3	1	2	1	1	5.0	1.6
200	5	5	5	5	3	3	3	3	4	3	2	2	1	1	1	1	1	5.0	1.0

for many of the items extend over the 1 to 5 range. Some of the items are rated a bit lower on the average than the other items, but the major feature is the variety of ratings for each item. Having an array of ratings such as this, one may inquire whether the differences are random over the raters or whether there may be systematic agreement for groups of raters. To answer this kind of inquiry, we perform a type of obverse factor analysis using each judge as a variable.

The first step in the factor analysis was to compute the sum of squares of ratings for each judge and the sum of products between each pair of judges. Mean square and mean cross-product values were obtained by dividing the sums by the number of items, 225.

Before considering the factor analysis, it will be of interest to consider the matrix of intercorrelations between the judges in Table 2. Note that the judges 5, 14, and 3 have high intercorrelations. Note also that the ratings of these judges correlate negatively with the last five judges, 12, 13, 15, 17, and 11, while these five judges intercorrelate positively within their group. The judges in the middle group have only moderate correlations with no distinctive patterning. This inspection tends to indicate the existence of two schools of thought as to what kinds of items are most relevant for the evaluation of social studies abilities at the end of secondary school. The factor analysis to be discussed will bring this observation out in more precise form.

Returning to the main thread of our analysis, the factor analysis of the matrix of mean cross-products, communality estimates were inserted in the diagonal cells of the matrix. The characteristic roots and vectors of the corrected matrix were found in order to obtain principal axes factors. There seemed to be two significant factors. The loadings of these two factors are presented in Table 3. Loadings on Factor I are approximately equal to the mean ratings for the judges. Loadings on Factor II range from .99 to —1.06 with the first three judges having positive loadings and the last five judges having negative loadings. The middle nine judges have near zero loadings on Factor II. Note that the judges have been listed in descending order of their loadings on Factor II.

Interpretation of the results is now to be considered. A rotation of axes could have been performed in the two-dimensional space defined by the two principal axes factors, but there did not seem to be any preferable location of the axes over the original ones. Remember that the loadings on this first factor approximated closely the mean ratings of the judges; consequently, we may in-

Table 2

Matrix of Intercorrelations

Decimal Points Are Omitted

	5	14	3	4	7	6	9	1	16	2	8	10	12	13	15	17	11	Mean	SD
5		66	69	21	42	20	32	25	19	15	18	03	−12	−26	−15	−28	−45	3.09	1.31
14	66		55	15	41	28	12	41	27	07	20	02	−20	−27	−02	−23	−35	3.17	1.12
3	69	55		17	37	19	27	33	27	14	31	−06	−08	00	−10	−14	−33	4.14	1.04
4	21	15	17		18	17	−20	00	11	09	04	05	01	−08	04	04	−14	4.45	1.01
7	42	41	37	18		18	25	17	46	20	16	11	−04	−02	24	25	−07	2.74	.57
6	20	28	19	17	18		06	24	08	16	24	05	10	05	28	11	−02	2.36	1.25
9	32	12	27	−20	25	06		19	15	11	08	07	21	10	08	06	−07	3.15	.66
1	25	41	33	00	17	24	19		22	14	39	13	10	01	29	11	09	3.14	.81
16	19	27	27	11	46	08	15	22		06	29	08	−03	−14	32	20	11	3.52	.77
2	15	07	14	09	20	16	11	14	06		12	08	24	08	18	21	05	3.12	.66
8	18	20	31	04	16	24	08	39	29	12		−03	17	04	29	19	30	2.48	1.07
10	03	02	−06	05	11	05	07	13	08	08	−03		21	22	23	22	08	2.68	.89
12	−12	−20	−08	01	−04	10	21	10	−03	24	17	21		29	39	35	33	3.47	.89
13	−26	−27	00	−08	−02	05	10	01	−14	08	04	22	29		19	24	24	2.58	1.17
15	−15	−02	−10	04	24	28	08	29	32	18	29	23	39	19		37	45	2.52	1.14
17	−28	−23	−14	04	25	11	06	11	20	21	19	22	35	24	37		53	2.39	1.38
11	−45	−35	−33	−14	−07	−02	−07	09	11	05	30	08	33	24	45	53		3.68	1.51

Table 3

Principal Axes Factor Matrix

Principal Axes Factor

Judges	FACTOR I	FACTOR II
5	3.13	.99
14	3.20	.74
3	4.17	.64
4	4.44	.17
7	2.76	.12
6	2.40	.11
9	3.15	.08
1	3.17	.08
16	3.54	.03
2	3.13	− .03
8	2.54	− .06
10	2.68	− .14
12	3.48	− .37
13	2.58	− .43
15	2.58	− .48
17	2.44	− .79
11	3.69	−1.06

terpret the first factor as a measure of general approval of the sample items. The magnitude of the loadings for this factor in relation to those for the second may be interpreted as indicating considerable agreement among the judges concerning the appropriateness of the items.

The second factor indicates a difference of opinion between two groups of judges. In order to interpret this difference in opinion, mean group ratings were found for each item as illustrated by the last two columns in Table 1. Inspection of these two columns reveals a distinct trend from top to bottom of the table for the ratings to increase for Group A (judges 5, 14, and 3) while the ratings for the other group tend to decrease. A scatterplot was made between these two sets of group ratings and is presented in Table 4. The correlation appears to be very low and is, in fact, negative. Most striking is the array of items rated 5 by all three judges in Group A. The other group of judges tended to rate these items down.

The item types and content that are rated high by one group and low by the other group are of interest. Those items rated high by Group B and low by the other group were the "information" items, calling for recognition of terms used in the social studies and the recall of fairly specific information. In comparison, the

Table 4

Scatterplot of Group Ratings

Ratings: Group B (12, 13, 15, 17, 11)	Ratings: Group A (5, 14, 3) 0.0–0.9	1.0–1.9	2.0–2.9	3.0–3.9	4.0–4.9	5.0	Total Frequency
5.0							
4.0–4.9			4	14	5		23
3.0–3.9		5	35	41	27	3	111
2.0–2.9		7	12	5	27	8	59
1.0–1.9		1	3	4	11	13	32
0.0–0.9							
Total Frequency		13	54	64	70	24	225

Mean Ratings:		
	Group A (5, 14, 3)	3.5
	Group B (12, 13, 15, 17, 11)	2.9

items rated high by Group A and low by Group B were the "self-contained" items, involving reasoning and interpretation from given material, such as reading comprehension of passages or interpretation of data in the social studies area. The items classified as "composite," requiring both knowledge and skill, which were actually most numerous in the sample, tended to show no such sharp differentiation between the groups.

As a further aid to interpretation of the opinions and factors, general comments and criticisms by the judges were collected.

From the item rating and comments, it seemed clear that Group A emphasized for both secondary school instruction and examining the gaining of problem-solving abilities. Group B, on the other hand, emphasized at this level the development of a facility to organize material and to express generalized conclusions effectively. This second group tends to feel that, if there is any value in the objective test, it is in measuring the knowledge which is a prerequisite to organizing and presenting thought.

The results of this study, we feel, tend to demonstrate that basic differences in value patterns can exist between groups of specialists in a content field and that the application of factor analysis to ratings of test questions can identify the groups and clarify the nature of the differences. The collecting of general comments can serve to insure that ideas not included implicitly in the sample of items are brought to the attention of the investigator. In some cases, points raised in the comments may indicate the necessity

for augmenting the pool with additional kinds of questions and repeating the analysis. The analysis can also provide an index of the magnitude of the differences thus identified. We may suggest that, in order for a generally acceptable test to be constructed, it may be necessary to provide part scores so that groups with differing points of view may utilize different weights reflecting their special value patterns when evaluating examinees. Thus, the methodology outlined is adequate to indicate those cases when high general content validity is possible as well as to indicate the existence and nature of ambiguity and controversy relevant to the content validity of a test.

Session II

Theme:
Use of
Achievement Tests in
Award of Course Credit

JACK N. ARBOLINO,
Advanced Placement Program, College Entrance Examination Board

Credit and the Advanced Placement Program

THE COLLEGE ENTRANCE EXAMINATION BOARD'S Advanced Placement Program has passed its probationary period. Growth has been sound and steady even though not spectacular. In May 1961, 13,283 candidates from 1,126 schools took 17,603 examinations and their scores were sent to 617 colleges. In 1956, the first year the Board took responsibility for the Program, 1,229 candidates from 104 schools took 2,199 examinations and their scores went to 130 colleges. In Dr. Conant's latest book, *Slums and Suburbs*, he says: "The success of the Advanced Placement Program in the last few years is one of the most encouraging signs of real improvement in our educational system." He also says: "...every high school ought to strive to provide the opportunity for advanced placement in at least one subject, no matter how few candidates there may be."

Frank O. Copley, in his book, *The American High School and the Talented Student*,* says of the Advanced Placement Program: "Few faults, if any, have been found in it by students, teachers, or administrators, whether on the secondary or college level; praise and satisfaction have been well-nigh universal." President Keeney of Brown has described the Program as "the most promising thing that has happened in American education in years." And Paul Roundy of Western Reserve Academy sums up his feelings about it with: "We have only our joys to celebrate."

Things are not that good. There are places where they have never heard of Advanced Placement. That is not serious; but there are other places—and they are nearby—where people wish it would just toddle off, with or without its powerful, nonprofit father. One of the most common charges against the Program is that it accents credit and by overemphasizing "testmanship" may

* Ann Arbor: The University of Michigan Press, 1961.

lead to the breeding of a race of what Henry Bragdon calls "academic spigots." The charge of spigotry can be rather easily refuted, but the problems of credit and testing had better not be denied if honesty and understanding are important.

Only about one and one-half freshmen out of a hundred are Advanced Placement candidates and the Program is still young. So, although an occasional sweet bird of youth may fly by the nets of regimentation, the nets are still there. Academic bookkeeping remains an integral part of American education, and even though no committee on instruction would admit to giving splendor to art by numbers, I have never known a registrar who did not count. Some of them, to be fair, will take the cash and let the credit go, but almost all of them, accurately and inexorably, count. The Program, then, must conform. It must, despite the finer aspirations of its originators, be concerned with such mundane matters as credits, points, and, to put it as crassly as possible, payoff. Study *is* its own reward, but because in a fairly rigid national system any innovation must fall heir to the ills as well as the attributes of that system, Advanced Placement must deal in the common currency.

Before describing some of the procedures used by the colleges in granting credit, it may be interesting to discuss attitudes toward the granting of credit by examination. Indeed it may be true that because policies and practices are changing quickly, attitudes are more important and revealing than statistics. To illustrate attitudes toward the Program and specifically toward the granting of credit, three stories may have value.

One of my first talks about Advanced Placement was addressed to the Catholic Education Association of an eastern state. During the question period one of the sisters observed that the Program held a great deal of promise. Then a priest, rather archly I thought, said, "I presume the sister is from a secondary school. Would she be in favor of allowing a student from elementary school to skip some of *her* work?" The sister rose and said, "Is the father addressing me?" and when he nodded, she said, "Yes," and sat down. The brevity of her answer left a gap. Then the priest got up and said—and this was meant to be the real clincher—"Would she be in favor of doing the same thing with postulants and novitiates?" Again she rose and said, "Is the father addressing me?" He nodded and again she said "Yes" and sat down. I had learned something.

Shortly after, at the University of Chicago, I was told that although the University favored the Program, an Advanced Placement candidate who had successfully completed one of our exam-

inations would be considered for credit only if he passed one of Chicago's placement tests as well. I returned to the office feeling wronged if not wounded. After I told Mr. Bowles of my experience, he thought a moment and said, "Jack, there is nothing one socialist hates so much as another socialist."

And finally the third story: More than a year ago I was invited to a meeting of the "in-group" of Advanced Placement colleges. These were colleges that had received the largest numbers of Advanced Placement students. The meeting was informal, there was no agenda, the purpose was merely to discuss common problems and experiences pertaining to the Advanced Placement Program and the matter of granting credit. After the Princeton representative spoke, a dean from Northwestern, emboldened perhaps by the fact that he was a Princeton alumnus, said, "Come now, you don't *really* give much credit on the basis of Advanced Placement Examinations, do you?" The Princeton spokesman said, "Indeed we do. Obviously you haven't read our latest statement. As a matter of fact, it is *possible* at Princeton to get as much as two years' credit on the basis of Advanced Placement." The Northwestern dean was silenced, but the Harvard man said, "See your ten and raise you five."

The leadership and vision of a small number of colleges are among the powerful forces working to shake the complacency of a larger number that have yet to awaken to the benefits of Advanced Placement. It is true that the few have reaped a disproportionate share of the rewards but, what is more significant, is that they have taken a step to establish standards in which America can take some pride. Those that have not yet moved might well heed Satchel Paige's admonition: "Don't look back; something may be gaining on you." For the second consecutive year, a National Education Association survey reveals that the secondary schools will emphasize college-level work. Therefore it would seem only prudent for colleges which have not taken a stand on credit by examination to drop some of their premises or at least start looking around.

John Valley of the Educational Testing Service, one of those rare men who can exert tremendous effort without making a sound, published during 1959 and 1960, five articles dealing with the award of placement or credit by subject.* The candidates who

* English: *The English Journal*, Vol. 48 (1959), pp. 398-401; Latin: *Classical Journal*, Vol. 55 (1960), pp. 215-16; modern foreign languages: *The Modern Language Journal*, Vol. 63 (1959), pp. 261-63; history: *Social Education*, Vol. 23 (1959), pp. 330-32; sciences: *The Science Teacher*, Vol. 26 (1959), pp. 399-402.

offered mathematics fared best. For a general appraisal, not by subject, a study in 1960 which covered 83 per cent of the students who took Advanced Placement Examinations in 1959 indicated that the majority received advanced placement or credit or both when entering college. The figure was 62 per cent. A second finding considered especially significant by the Advanced Placement Committee was the fact that the colleges which substantial numbers of Advanced Placement candidates have already entered, are the most liberal in awarding placement and credit. That figure was 82 per cent. It seems clear then that with experience—more specifically, experience with candidates—college policies become more liberal.

The policies vary. Some, like the College of Literature, Science, and the Arts at the University of Michigan, award credit for examination scores of 3, 4, and 5 and refer scores of 2 to the departments concerned for individual evaluation. Other colleges grant placement only and still others, although they are now hard to find, give neither placement nor credit. It is not easy to summarize policies toward the Program because not only do colleges within universities react differently, but even departments within colleges sometimes differ in their reactions.

A year ago 457 colleges had indicated that they granted placement or credit or both. Doubtless that number has increased, and among those colleges making up this first number there is no doubt that the policies of many have changed. It is an exaggeration to say that certain Advanced Placement policies are like missiles—if they work they are obsolete. It is dangerous, however, to cite a policy that is more than a year old. It can also be misleading to cite policies or even statistics as being up to date on the percentage of candidates who receive placement or credit. Students and secondary schools must consult the individual colleges for specific answers.

The ultimate justification for awarding credit to successful Advanced Placement students lies in their performance in sophomore courses. Every report received by the College Board indicates that these students do at least as well as, and frequently better than, the regular sophomores.

There are, of course, some problems ahead for the college that takes on the task of properly placing and crediting these students. However, the problems are no more difficult to solve than those accepted by the secondary school when it institutes Advanced Placement courses. And the obligation of the college is at least as great.

One of the objectives of the Program is to effect indirectly the proper placement and accreditation of successful candidates. This frequently calls for the conversion of colleges and it entails politics as well as pedagogy. Had we but world enough and time, the pulls and pushes would always be gentle.

Colleges will continue to take whatever action they will with candidates. That is their prerogative. But students and schools have a privilege too and it is just as proper. One can hope that the colleges that condescend, or do not deign, or dissemble by looping the problem with rolls of rhetoric, will be found out. That should not be too hard. The strongest resistance comes from guerrillas--dedicated, mostly scattered, some even isolated on their own campuses: the head of freshman English who would make Milton register for composition; the new key man in French who will not even concede that translation to English may be a useful testing device; the mathematician who, deploring the way calculus is taught in our best colleges, says: "Why bring it down into the schools?" These are our dearest enemies and they are tough and able.

There is another group, not enemies but both less and more. They treasure learning, dwell in airy castles, and crest the academic world. Students, schools, and programs swirl beneath them. They are above the elements they live in. With them social responsibility and academic detachment may never be reconciled, but they had better be. And there is hope, for truly to treasure learning is to transmit it. A castle or even a palace of art is barren if it is lonely.

The Examining Program at Antioch College

RUTH D. CHURCHILL,
Office of Testing and Evaluation, Antioch College

THE EXAMINING program at Antioch College may be taken as an example of the way in which such a program can be integrated with, and influenced by, the educational program of a college. While it is no novelty for a large university to have comprehensive examinations, it may seem less possible for a small institution. On the other hand, what the small institution lacks in terms of resources—financial, technical, and otherwise—may be compensated for by the intimacy of the relationship among all parts of the curriculum.

Antioch's present examining program came into being twenty-one years ago as part of its program of general education. It was instituted after a six-year study of achievement in general education courses. At that time a set of fourteen courses was required of all students. The purpose of the testing program was two-fold: to set common minimum standards, but at the same time to allow for individual differences among students. The fourteen courses were organized into five area examinations which all students had to pass and which were conceived of as integrated wholes, not as series of course examinations. Those students who earned high scores were allowed to waive appropriate courses—as many as nine of the fourteen—and to earn credit by examination.

These examinations did not stand by themselves. They were preceded by placement examinations to aid students in planning programs. They were followed by a series of examinations in the senior year designed to evaluate the student's college experience. The senior examinations comprised a Terminal Integrating Examination in general education consisting of the Graduate Record Examinations Area Tests, a test in critical thinking, and a long essay; the field comprehensive examination, which included the advanced test of the Graduate Record Examinations; and a senior paper, in which the student reviewed his educational history.

This brief glimpse of the examining program as it was organized from 1941 to 1957 hints at many of the underlying purposes: recognition of input and output variables for purposes of both individual and institutional guidance; adaptation to individual differences in a framework which emphasized common standards; and use of examinations as a unifying and integrating factor in the curriculum. Five years ago the faculty of Antioch College, after two years of discussion and planning, approved a new general education program and a new program of examinations. The fundamental role of tests remains unchanged although the practice has responded to the needs of the new curriculum and to the new pressures and values emerging in higher education.

The new program attempts to place major responsibility on students for planning and carrying out a meaningful program of general education. At the same time it tries to give him assistance. A syllabus worked out by the faculty indicates the scope of general education in terms both of content and of objectives ideally to be attained. Within the framework of this syllabus, students plan their programs and faculty teach such courses as seem to them best suited to achieve the aims; no course exists which students must take or faculty must teach. In addition, courses are grouped in three areas and at two levels, followed by a general education seminar in the senior year. The curricular requirements for students are that they select two courses in each area at each level and that they complete the first level in an area before they begin the second.

The role of the examinations involves both guidance and evaluation of achievement. As soon as the student arrives on campus, he takes a series of five comprehensive examinations. Two of these are skills examinations in English and mathematics, setting minimum standards which students are expected to meet at entrance. If a student does not meet these standards, remedial help (outside courses) is available. The requirement must be met by the end of the second year, since the skills involved are obviously basic to college work.

The other three entrance tests are area examinations in the physical sciences, the social sciences, and the humanities. These are the examinations which set the standards for the first level in the program. They are given at entrance to college for their value in guidance—these tests identify three groups of students with different needs who will follow different general education programs.

First, there are the students who fail to meet the minimum

requirements in one of the three given areas. Since scores relating to disciplines are available for each area examination, such students are not only told of their general need for background but also given specific information about fields in which they need courses. While only two courses are required in an area, the effect of this warning, combined with the necessity of passing the examination before the end of the third year, is to encourage students to do additional work, either in the form of courses or of independent study.

A second group of students meets the minimum standards. These students are relieved of further concern with the examination; at the same time they are given information about their strengths and weaknesses, which they can use in choosing courses. No pressure is exerted on them to choose courses either in fields in which they are well prepared or in those in which they lack background.

The final group of freshmen identified consists of those who do exceptionally well on the area tests. Such students earn credit by examination for Level I courses in the appropriate area. All six first-level courses may be waived and credit earned.

In recent years there has been much concern for the stimulation and intellectual enrichment of able high school students and for the better articulation of high schools and colleges. The Advanced Placement Program of the College Board has been one answer to this problem. This year at Antioch 36 students out of an entering class of approximately 475 took 51 Advanced Placement courses and received college credit for half. Almost all those granted credit, and some of those who were not, earned credit on the area examinations. (No one was given double credit—that is, credit for both course and examination.) The point is not that these programs overlap, yielding evidence of mutual validity and causing minor administrative problems, but that the Antioch program extends the concept of advanced placement to all students. And such a program is needed—in visiting high schools I have noticed the existence of a variety of honors programs not formally given for college credit but markedly enriching background in many areas.

I have stressed the guidance aspect of the comprehensive examinations, but it must not be forgotten that all students must pass these examinations to progress in their general education program. And the examinations do cover a broader scope than any combination of two courses which a student may take. Therefore, they hold students to a breadth of preparation in each area and expect some independent effort on the part of the student to attain that breadth.

In summary, the program at Level I sets minimum standards,

gives students guidance in planning courses, and permits qualified students to begin their college program at the highest level of which they are capable. Presently, the skills and the Level I examinations are constructed by the faculty. (Basic policies are set by a faculty committee.) The advantages are that the examinations can be directly related to the objectives of the college and that the faculty know the examinations. The disadvantages are lack of technical facilities, unevenness in faculty ability to work on examinations, and the amount of time demanded by test construction. Perhaps the final product is never a polished examination. Wherever external examinations whose objectives and standards coincide with those set by the faculty are available, they are used. For example, the faculty committee at Antioch is considering the College Board English Composition Test and Writing Sample for use as the skills examination.

Two other features of entrance testing at Antioch should not be neglected even though they are outside the structure of the achievement program. The first is the use of College Board achievement examinations for advanced placement in field courses in foreign languages, mathematics, and chemistry. Generally these tests have been taken before students arrive at Antioch. The second is that information about aptitudes, interests, and general background of students is also obtained. These data are available to students, their advisers, and special advising personnel for help in choosing major fields and vocations. They are also available to the college for planning and research. Since these data are transmitted to students and faculty, personality scales are not included—these are given only for special research projects.

Let us return to the achievement examining program. During his first two to three years each student pursues the first level of his general education program. As he finishes Level I within an area, he is free to proceed to Level II in that area. At this second level no courses are waived by examination, and the examination itself changes in character. A single examination covering all areas is given at Level II. One part is composed of the Area Tests of the Graduate Record Examinations; the other is a locally constructed test involving logic and reasoning with data. The Graduate Record Examinations have the merit of being well integrated, interesting presentations, although they may be somewhat too easy—there is no evidence that they are more difficult than the Level I examinations. They have the advantage of providing, in a sense, an outside examiner in the general education area. Using these examinations, it will be possible this year to compare the last complete

class graduated under the previous general education program with the first complete one under the present program. Moreover, it will be possible to discover whether or not differences in availability and choice of courses and earlier timing of the examination --the Level II examination is taken when courses at this level are substantially complete, about a year earlier than under the old plan—produce differences in final level.

One other examining device is employed at Level II—a check on writing. Advisers and instructors are queried about the day-to-day writing of upperclass students. Students whose writing appears unsatisfactory are referred to the literature department which reads recent papers and provides help and consultation.

In the senior year the examining responsibilities pass from the general to the field education program. However, the senior general education seminar is an outgrowth of the old Terminal Integrating Examination. Students and faculty valued the discussion which preceded the writing of the essay for that examination, and in planning the new program provision was made for senior seminars drawing students from all fields to study topics of general importance. In the major field, students have departmental seminars and field comprehensive examinations including the Advanced Tests of the Graduate Record Examinations.

This description of examining at Antioch has reflected some of the principles, purposes, practices, and problems of a complex program. Problems are manifold and change continuous. At present, for example, we are concerned with the skills examinations and with the possibility of adding another which tests ability in a modern foreign language. The problem of how to set standards on achievement examinations and how to maintain them from year to year is perennial. The need for research haunts us; at the same time we are conscious of imposing too many such experimental demands on students. Without much more detail we cannot begin to portray the delights and the anxieties inherent in giving comprehensive examinations as we do at Antioch in close contact with the "consumers." Let us close by saying, "The unexamined life is not worth living."

Examinations and Instruction by Television

CLIFFORD G. ERICKSON,
Chicago City Junior College

THE USE of television as an aid to instruction affords many opportunities for asking fundamental questions about the teaching-learning process and for finding new solutions to old problems. It is the purpose of this paper to suggest the unique role of televised instruction in the development and use of evaluation instruments. These remarks will be based on the assumption that the use of television as an instructional medium is accompanied by basic adjustments in the way of organizing teachers and pupils, in the assignment of teacher time, and in the use of specialized non-teaching personnel.

The best qualified and most able teachers are usually selected for the preparation and presentation of television instruction. This places the teacher leadership in the development of evaluation instruments in the hands of those teachers most capable of providing proper direction for this work. These are teachers most likely to be able to utilize the assistance of specialists in curriculum, in learning media and materials, and in evaluation. They very often have or can develop the ability to direct a cooperative situation in which many teachers can participate in the important process of defining objectives, selecting materials, developing evaluation instruments, and interpreting results.

The use of television should and usually does involve a sufficient expenditure of dollars in equipment and technical aid to make administrators and controlling boards aware of the need for similar investment in specialized assistance in curriculum, learning materials, and evaluation. There is, moreover, the recognition of the need for the assignment of sufficient time to the television teacher and his colleagues to enable them to properly shape the objectives and content of the program of instruction, to utilize the aid of specialized resource people, and also to make use of available knowledge about the teaching-learning process.

Whether the instruction be total instruction at the college level or supporting instruction at the elementary or secondary level, the television teacher finds himself either directly in a cooperative teaching situation, or he must make careful preparation for supporting teachers more remotely related to him. This new teaching-learning situation enhances the opportunities for communication between colleagues and the refinement of ideas leading ultimately to valid and reliable evaluation instruments. This kind of team effort is not easily achieved in conventional instruction because teacher time is not traditionally organized to allow for this activity.

Television as a means of multiplying the effective use of a good teacher is, by the same token, a way of multiplying the use of carefully prepared evaluation instruments which are validly related to the given learning situation. More teachers can have the experience of building good test items, analyzing results, and improving the test materials. More pupils fall into an evaluation program based on sound principles and thorough preparation.

All of this provides a situation and climate for valuable research in curriculum development, in professional growth for the faculty, in improvement of instruction, and in interpretation of the educational program to the institutions participating.

The foregoing statements are illustrated by our experiences at TV College—an integral part of the educational service of the Chicago City Junior College. Since 1956 we have been presenting college courses for credit via open-circuit television to the residents of the Chicago area. Our teachers are carefully selected from among those who have demonstrated a high level of teaching skill, sound scholarship, professional maturity, and an ability to work in a cooperative situation. These teachers are given the equivalent of eight weeks of full-time work at full pay to prepare for the presentation of the telecourse, and work full time presenting two 45-minute broadcasts per week during the course.

The television teacher's first deadline is the preparation of a statement of four or five main course objectives with an explanatory paragraph on each. Explicit and concrete statements of course objectives are desired in order that they may serve as a guide for selecting and organizing the learning experiences, for preparation of assignments for students, and also for devising appropriate means of evaluating student learning.

This opportunity to prepare carefully a set of course objectives with the help and guidance of specialists is one of the unique values of our TV College program. Teachers of conventional courses seldom find time to prepare explicit formulations of course

objectives. Their courses tend to develop as they teach them, following a vague outline of implicit objectives, and often take shape from available textbooks or materials. Even multi-section courses developed by committees often evolve similarly, representing the results of collective past practice rather than the results of extensive educational planning. Since college teachers are usually specialists in content both by training and by interest, their courses tend to develop around details of content rather than around principles derived from broader behavioral goals.

With time and expert assistance available to them, television teachers develop considerable skill in formulating statements of objectives. Early drafts of such statements are often content-oriented, for example, a listing of content-unit headings—stating what the college teacher was going to do rather than what behavioral changes he wants to bring about in his students—or, a statement so general that it means different things to each college teacher. They acquire the ability to formulate course objectives in terms of behavioral changes as well as learning content and in terms of attitudes, appreciations, and abilities as well as knowledge or recall of information. Many teachers become aware of the importance of developing critical thinking as a major course objective. As time goes on they are able to expand these statements in terms specific enough to indicate appropriate procedures to section teachers and control-class teachers, as well as aiding the TV teacher himself in selecting and organizing learning experiences and in constructing pre-test, midterm, and final examinations.

The television teacher proceeds to use these objectives as a basis for the outline of telecast topics, the selection of textbooks and other learning materials, and the selection of audio-visual materials for his telecasts. He then undertakes the preparation of the teleclass study guide for integrating the televised instruction, textbooks, and collateral readings and assignments into a unified program of learning for the TV student.

Finally the television teacher begins preparation of evaluation instruments for his course, usually two midterm examinations and one final examination. Expert advice and assistance from the Coordinator of Research and from the Division of Student Examinations enables the television teacher to acquire considerable skill in test construction. He learns to produce tests with items distributed evenly over the major objectives of the course, with a variety of types of objective questions, and with simple, clear instructions for responses. Essay tests receive clear criteria for evaluation. He learns to use item-option analyses to determine the dif-

ficulty and discrimination of items. When possible, he administers the test to trial groups to make item analyses and to improve items showing poor difficulty and discrimination. All of this provides teachers opportunities for their own professional improvement—of benefit not only to them but also to their colleagues in subsequent classroom teaching.

Much of the foregoing implies directly or indirectly that television instruction has brought to the Chicago City Junior College new opportunities for upgrading courses, improving faculty competence, and reaching students who might not otherwise be reached.

Using examinations developed in this setting, how well does television succeed with this kind of instruction? In our third comprehensive report, *Chicago's TV College*, we summarize a host of achievement comparisons made in some fifty different televised college courses:

> "In the first year, achievement comparisons were made for three groups of students: the at-home television students receiving 30-minute telecasts, the conventional classroom students receiving 50-minute conventional lessons, and the classroom students receiving thirty minutes by telecast and twenty minutes of follow-up classroom instruction. In all subjects taught—English, biology, social science, political science and mathematics—the at-home television group showed highest achievement in raw scores. Equating the groups by matched pairs of covariance analysis the differences in biology were significant at the five per cent level favoring the home television student. In the other subjects, there were no significant differences among the groups.
>
> "In the second year, achievement comparisons were made primarily between TV-at-home and evening adult classroom students. In accounting, English, social science, psychology, and shorthand there were no significant differences between the control and TV groups. Significant differences were found, favoring the TV group, in humanities, biology and, in one instance, physical science. These differences may be explained in part by the type of course, the population involved, and differences in instructors.
>
> "In the third year, variations were found in the comparisons of TV-at-home, TV-in-class, and control classes. The conventional control groups in Physical Science 101 and Humanities 202 were favored over TV-in-class groups at the five per cent and one per cent levels respectively. In Child Psychology 207,

Fundamentals of Speech 141, and Mathematics 103-College Algebra, TV-at-home groups were favored at the five per cent, one per cent, and one per cent levels respectively in comparison with the control groups. In Social Science 102, Humanities 201, and Speech 141, TV-in-class and conventional day groups did not differ significantly."

At the close of our first three years of operation, supported in part by a grant of $500,000 from the Fund for the Advancement of Education, a panel of distinguished educators was invited to study our operation and make an evaluation. This group included Dr. Frederick L. Hovde, President, Purdue University; Dr. Henry Chauncey, President, Educational Testing Service; Dr. Norman Burns, Secretary, Commission on Colleges and Universities, North Central Association of Colleges and Secondary Schools; and Dr. Wilbur Schramm, Director, Institute for Communication Research, Stanford University.

In its summary of findings it included the following statements:

"Courses at the junior college level can be taught effectively to a home audience by television. The results on this point were most impressive and convincing. Indeed, in the few cases where there were significant differences between the performance of home TV students and classroom face-to-face students, the differences were more often in favor of the TV students than the others.

"Evidence of the high motivation of the TV students and the welcome given the TV courses is the fact that about 65 per cent of the television students finish their courses and take the final examinations. This completion rate is quite remarkable when compared with other forms of adult education for credit.

"The experiment did not produce evidence to show that teenage college students can be effectively taught exclusively by TV. It may be that these younger students need more direct contacts with the teacher. However, it is now quite clear that TV can be used as the main channel for teaching the more mature and more highly motivated students who are chiefly attracted to the home classes by television. Even in the case of these students, however, it is evident that completion and high quality work can be encouraged by mail-in assignments, trial tests, and the possibility of face-to-face conferences and telephone conference periods.

"The experiment appears to have generated a healthful ferment throughout the junior college system in Chicago. The early fears of teachers that they might be superseded by a picture

tube appear to have been mostly done away with. In their place has apparently come a new interest in stating clear course objectives, in using TV study guides in classrooms as well as home, in making better examinations, on the pattern of the examinations for TV students, and in the experience of television teaching. We note with interest that whereas at first it was difficult to get volunteers for TV teaching, now the applications greatly exceed the openings. . ."

Perhaps those brief remarks about our experience in the Chicago City Junior College have borne out my initial statement that the use of television can provide unique opportunities for improving an instructional program, for upgrading methods and materials for evaluation of the program, for measuring student achievement, and for awarding grades.

Credit and Waiver Examinations at Michigan State University

WILLARD G. WARRINGTON,
Office of Evaluation Services, Michigan State University

THE PRECEDING papers have been primarily concerned with the awarding of credit by examination under a variety of conditions. I wish to discuss briefly a similar program at Michigan State University, and I also wish to introduce the proposition of using achievement examinations to waive required courses without credit.

Michigan State University has long had a policy by which students could gain credit for any course by passing an appropriate examination. In practice, however, this policy is rarely implemented in most of the colleges of the University except in very atypical instances.

But, within the University College, a very different philosophy prevails regarding credit by examination. For those of you who may not be familiar with the Michigan State program, I should explain that the University College is the administrative unit for our program of general education. At Michigan State, a total of 45 quarter hours of general education is required of all students who matriculate in the University. These general education credits are normally earned by enrolling in the three terms of each of our four general education courses. One of the unique characteristics of the University College program is that all students registered in a particular University College course must take a common term-end examination that accounts for one-half of their final grade in the course. These term-end examinations are prepared and administered under the direction of professional examiners in the Office of Evaluation Services.

The initial planners and organizers of the University College recognized that a requirement that all students take all required courses in exactly the same prescribed manner would be inconsistent with what we know about individual differences and the heterogeneity of college students. Such rigid requirements are

particularly inconsistent with the philosophy underlying a general education program. For, if there are fundamental areas of knowledge and experience which should be the common possession of all educated men and women, and we firmly believe that there are, then it is to be expected that this knowledge and experience could be obtained in a variety of fashions and under a variety of conditions. In the University College, therefore, acceleration by earning credits by examination without conventional enrollment in courses is an integral part of the total general education program.

In practice, this acceleration privilege has operated as follows. Students considered qualified are allowed to take the same term-end examination that is taken by students registered in the course, without themselves being enrolled in the course. If they earn an "A" or "B" on this examination, full academic credit is awarded and recorded on their official transcripts. While there has been some variation over time and among departments, the usual practice is to enroll in the first term of a three-term course. Students demonstrating exceptional performance in this first term of the course, that is, usually having an "A" at midterm, may be granted special permission to take the term-end examination for the next two terms of the course without enrollment. Similarly, students having taken two terms of the three-term sequence and having an "A-B" combination in the two terms may be granted special permission to take the term-end examination for the third term without enrollment in the course. Other criteria have been used in this selection process, but the primary criterion has always been performance in the initial term or terms of the course.

It should be kept in mind that this special-permission student takes the examination with the much larger group of students actually enrolled in the course. Credit is awarded when this student demonstrates a level of competency equal to or better than that awarded a "B" for students enrolled in the course. Failure to demonstrate this level of competency means that the student must enroll in the course in the conventional manner. This, then, is a very brief description of our acceleration program as it presently functions. I would now like to discuss a few questions that might be asked concerning this program.

Do students utilize this acceleration privilege? The answer is unquestionably "yes," in that acceleration occurs at a much greater rate in the University College than in the other colleges of the University. For instance, during a four-year period, between 1949 and 1953, one study indicated that approximately 4.5 per cent of

the credits earned in the then Basic College were earned by acceleration. But, since that time, there has developed a tendency downward in the proportion of credits earned by acceleration, reaching a low of approximately of 1.5 per cent in the academic year 1958–59. This trend has now shown a slight reversal, since during the academic year 1960–61 approximately 1.7 per cent of the credits earned in the University College were earned by acceleration. However, I should not create the impression that acceleration has become an insignificant element in our program, for during this past academic year, 1,326 special-permission examinations were attempted, resulting in the awarding of 3,252 hours of academic credit. While we have never attempted to justify our acceleration program in terms of economics, the sheer instructional manpower that would have been required to produce this many credits is surely not to be ignored. It is also generally agreed that increasing pressures from student enrollments, in combination with increased acceleration through the waiver system to be discussed later, will result in some continued increase in the rate of acceleration in the University College.

What kinds of students accelerate? First, let me say that not all students who are eligible choose to accelerate, and certainly not all who attempt to accelerate are successful. There is some evidence that attitudinal differences exist among those who accelerate and those who are eligible but do not. Specifically, our data suggest that those who accelerate are likely to be more aggressive, flexible, and generally at ease with the world than those who do not. Evidence also shows that reading ability is definitely related to success in acceleration attempts. This is certainly not surprising, since in reality our acceleration program is primarily an independent study program wherein the student acquires the knowledge to pass examinations by reading.

Is successful acceleration a predictor of future success? This has been a source of considerable controversy since many faculty members would argue that passing an examination without being in the course is different from passing an examination after being in the course. Here, I can only report our findings—namely, that students who accelerated in the University College courses did better work in advanced courses in the same area than those who were eligible but did not accelerate. Thus acceleration appears to be in no way a detriment to achievement in closely related but more advanced courses. It should be pointed out that the fact that some eligible students do not choose to accelerate may reflect undue concern for grades, since these students apparently choose to

enroll in the course expecting to make an "A" rather than taking a chance on accelerating and possibly earning only a "B."

The question as to how students prepare for acceleration examinations often comes up. Our studies would indicate that such students prepare in pretty much the same way as students who are actually taking the course. That is, they read the material, they study the assignments, and often they acquire notes from students taking the course and study them.

In summary, then, it would seem that our acceleration program is reasonably effective. Some would like to see the privilege utilized more by students; others would prefer to see fewer credits granted by acceleration. But in either case the available evidence would suggest that students do not seem to be penalized by the system and that, in fact, some students profit greatly from it.

I would now like to discuss briefly the waiver examination system that has only recently been introduced within our University College program. This waiver examination system has had its genesis from several origins. One consideration was that under our present acceleration system students normally are not allowed to accelerate the first term of the three-term courses. But, if students can accelerate in two terms, why not all three? Another problem arose when our examination system was changed from the single one year's comprehensive examination to the three separate term-end examinations. Under the present system, the student who wishes to attempt to accelerate in two terms of a course must take three two-hour examinations in a solid six-hour block of time because of our single examination administration. In practice, most students take two examinations and defer the third until the end of the next term, but in this case, they must also defer knowing whether they will be successful in this acceleration attempt. In addition to these two specific issues, continual opinion has been expressed that we should be more concerned about breadth and depth of educational experiences rather than an accelerated educational experience. The advocates of this position argue that students who by background experiences or independent studying can satisfy the general education requirements should be encouraged to take either more specialized courses or courses designed to increase their total scope of knowledge rather than just to accumulate acceleration credits. Thus, they recommend waiving required courses upon demonstration of competency, but without credit.

And a fourth consideration, quite unrelated to the others, involved the problem of the transfer student in our general educa-

tion program. For over the years, we find ourselves in the almost embarrassing situation of treating our own students by one standard with respect to general education requirements and transfer students by another. Specifically, we say that our own students can only meet the general education requirements by completing the University College program, whereas students transferring from another institution are often given complete or partial general education credit for having taken related introductory survey-type courses. It seemed to us that an argument could be made that such transfer students must demonstrate a minimum level of competency before the general education requirements should be waived. This is very tenuous ground, I realize, and we are treading very slowly and gingerly in this area.

From these considerations, our embryonic waiver examination system has developed, which represents a partial compromise between some of the conflicting issues that are involved. This waiver system may be described as follows: We have constructed one-hour examinations for each of the three terms of the four University College courses. These examinations are simply abbreviated term-end examinations and are administered one time each term during the registration period. The items in the waiver exams have been taken from previously used term-end examinations, the criteria for selection being that the item statistics were good and that the total test represents a good sampling of the course content.

After considerable discussion, two cutoff levels were established for each of the waiver examinations. The median level of performance of students actually enrolled in the particular course was taken to be the level required to waive the requirement that the course be taken. It was further decided that those students who performed on the waiver examination at a level equal to the "B+—A−" cutoff for students enrolled in the course should be allowed the privilege of attempting to earn credit for the course by taking the next scheduled regular term-end examination. If they earned an "A" or "B" on this term-end exam, they would be given credit for the course.

The decision as to what criteria should be used to determine who should be allowed to attempt the waiver examinations has taken an interesting turn. As the waiver system was being developed, rather elaborate specific qualifications for permission to take the waiver examinations were proposed. These involved evidence of high performance in high school and on the test battery given at the time of entrance to Michigan State, as well as high college performance for transfer students. But, before these cri-

teria were even used, it was decided such rigid standards were inconsistent with the purposes of the waiver examination system as we saw it. It seemed to us that the value of the waiver program would be enhanced if it were made very permissive. We hoped that students would see this alternative as both an academic opportunity and an academic challenge that could be the initial spur to increased individual initiative and effort. Consequently, the criterion for permission to take waiver examinations is simply that any student who feels himself qualified may attempt these examinations. At this point, we just do not know the consequences of this decision. The number of students attempting the examinations has increased each administration, but so far, our evidence suggests that most students are taking the examinations seriously. If subsequent data show this not to be the case, we may have to reinstate some selective criteria.

This system of waiver examinations went into effect for the first time in the fall of 1960. During the academic year, 1960–61, 251 students attempted 304 waiver examinations. Of these 304 attempts, 181 or 60 per cent resulted in successful waivers. Also, 59 student attempts, or nearly 20 per cent, were sufficiently high to allow the student the opportunity to attempt to earn credit by acceleration. While not all of these students have as yet made the acceleration attempt, those who have, have been earning credit in a proportion of about three out of four.

General reactions from both students and faculty have been quite favorable. All students attempting a waiver examination have completed a brief questionnaire in which we attempted to discover how students prepare for the waiver, how confident they are of passing the waiver, and why they are attempting the waiver. The data collected to date suggest that students are making a serious effort to prepare for these examinations. Most students read the text, class materials, and some of the outside readings. Only a very small proportion indicated that they depended on previous knowledge and experience and hence made no preparation.

We asked the simple question, "How would you estimate your chances of making a passing score?" We found that these students are relatively good predictors of their own success. When the students were classified into three groups, failed, waived, and waived with special permission, we found that the precentages checking reasonably certain or probable were 41, 65, and 83, respectively.

We also asked the purpose that motivated the waiver attempt and found the most frequent reason was so that the student could take other courses. Only a few indicated the purpose was to avoid

repetitious material, or to lighten loads, or to graduate sooner.

To summarize, it appears that in its initial stages, the waiver system is functioning reasonably well. Unquestionably, it has increased the flexibility of our general education requirements. Several problem areas have yet to be faced. We have made no attempt, for instance, to test transfer students for the purpose of validating transfer credits. Also, we will need to study rather carefully future course selection of students who successfully pass one or more waivers. We are assuming that students will be rational in their choices of substitute courses, but this will need to be verified. Also, since proportionately more new students attempted the waiver examinations this fall, we will need to be attuned to any impact that this waiver system may have on high school programs, but to date, we have no indication of trouble in this area.

In conclusion, then, I can report that in the University College at Michigan State University, we have no qualms about the validity of using examination performance to waive academic requirements or to award academic credit. However, it must be recognized that there are unique characteristics of our examination program, in particular, that would caution against immediate generalization to all other institutions.

Luncheon Address

Management of the Learning Environment

JOHN E. IVEY, JR.,
Midwest Program on Airborne Television Instruction

THERE ARE few university administrators and faculty members who have escaped the realization that the next decade will work great changes in higher education. To some, this realization is dimmer than to others. Responses to the forces of change range from downright aggressive opposition to open-armed welcome of the opportunity to modify those programs and practices that long since have become questionable.

A doubling of the number of students is in prospect. The biological, physical, and social sciences and the humanities are all swept up into world-wide change and ferment. Industry and government roam the academic environs snatching away top scholars and scientists with an appeal which sometimes blends patriotism and profit. The world around colleges and universities is changing much faster than the educational content and structure of most institutions. The struggle by higher institutions to provide national and international leadership becomes an out-of-breath romp to keep up with the times.

These strong pressures must be met by finding ways to reach even higher standards of institutional excellence in teaching, research and public service. Evidence is growing, however, that the financial requirements for most public and private institutions may not be met by their usual sources of revenue. New and improved educational operations are needed to merit financial increases from traditional sources of funds; or new financial sources must be found that will respond to old appeals.

To cope with these operational realities day after day is a grueling and often thankless exercise. Yet somehow, despite the pressures, higher education will not fade from the face of the earth. But will institutions allow the outside pressures to determine and limit their content and form? Or will their new content and form

lead in directions of their own choosing and be determined by methods of their own fashion?

Colleges and universities are crippled in the struggle to maintain control over the forces of change stirring from within and pressing from without. This arises, in part, because there exists no central body of educational theory or practice from which institutional planning and operations can be projected, tested, modified, and empirically verified and unified.

Several examples exist where a professional field such as engineering, or a basic discipline, such as biology, has begun to look at its own content and practices with a view toward self-improvement. Exciting excursions into learning theory and its application to improve teaching have been taken. The work of Professor Bruner at Harvard, the late Carl Hoveland at Yale, and Professor Skinner at Harvard, to mention only a few, are all important in this respect.

Yet the distance from these developments to formulating a crude theory of university administration and pressing it toward empirical testing is great. Such a framework for administrative guidance should encourage and accelerate the recasting and programming of the several academic and professional fields into some meaningful unity. What, then, are some of the ways in which educators, members of boards of trustees, legislatures, and the lay public might look at this peculiar social organism we call a college or university?

First of all, let us state the simple premise that a university is a learning environment. It is not a mere confederation of departments, schools, bureaus, and institutes contributing manpower to and supporting the same football team.

As a learning environment, the university provides students and faculty with value systems, traditional patterns of behavior, time and facilities which can catalyze or stifle the learning process. University management, teaching, research, and public service get much of their tone and vitality from the institution-wide forces operating in the learning environment. In turn, these various functions feed back to the university learning environment forces which either strengthen or undermine the values and behavior patterns the university attempts to nurture.

In this sense, the university learning environment is a unifying cultural system. Its operating divisions are subcultures that either contribute to, remain neutral, or weaken the unity and strength of the university cultural system.

If the parts are to draw strength from the whole, the various

units of the university cultural system, the faculty, administration, and trustees must work from a common set of educational purposes. They must be defined with sufficient preciseness that the purposes can be adequate guides for the performance of all institutional functions. The ultimate measurement of student development and faculty performance must be derived from these common purposes.

To the extent the definition of institutional purposes is cloudy or misunderstood, the university as a learning environment faces tremendous operational difficulties. This self-evident truth lies at the base of much stress in university administration and faculty effort.

Within the university learning environment, there must exist accepted limitations of effort. No institution can provide programs in all fields of study. Nor can it achieve the same degree of excellence in the fields to which it commits itself. Moreover, it cannot be expected to support indefinitely some fields that may have been left behind by the changing state of knowledge.

Each institution then has a basic problem of continually reviewing the epistemological structure it maintains. This requires a continual process of regrouping fields of study to reflect their emerging inter-relationships, the phasing out of certain courses and programs, the adding of new programs as resources for adequate support become available. Above all, this process should be kept consistent with institutional purposes, and within bounds of institutional financial capability for pursuing excellence in its range of commitments.

Within a given epistemological structure, what kind of curricular structure is required to give students adequate access to appropriate institutional scholarly and scientific resources? What kinds of behavioral capabilities are desired in the student for his optimal growth in each of the curriculum areas? Here is where broad institutional purposes, the precise academic goals of institutional units, the capability of institutional financial structure, and the performance expected of students have to mesh firmly.

The precise definition of outcomes in the mastery of each program area identifies for the student what the university expects of him. There can then be a meaningful exploration of how the growth process is to be guided, how student progress is going to be measured. This is the point where learning theory and curricular design converge in the conception of a unified university learning environment which will maximize student growth in each field of study.

Given certain results (defined in behavioral terms) necessary for student mastery of certain areas in biology, how can the student most expeditiously develop them? We must assume that, on the one hand, each field of study has characteristics that influence the design of learning experiences for students. These are inherent in the structure and content of the particular academic field. On the other hand, there are certain inherent and acquired characteristics in the human learning process that affect the manner in which the materials of a given field are introduced to and explored by the student.

The programming of the various academic fields within the university learning environment is a great uncrossed frontier. Due respect for the intrinsic integrity of an academic field, and the theory of learning which best governs its mode of exposition for student mastery, are at the root of fashioning the components of the university learning environment.

What are the appropriate roles of the student, professor, administration, and trustees in developing and operating the learning environment? What skills of learning does the student need in order to utilize fully the learning opportunities and tools accessible to him? How can the university assure that deficiencies in learning skills can be corrected in students before they become a critical deterrent to his growth in the institution?

These questions may have different answers, depending on variations in institutional purposes and modes of operation. Even so, looking across the total institutional endeavor, one must ask what administrative structure is required to operate the total learning environment? If each curriculum area is to be programmed, how is the developmental work to be accomplished and how are the different learning experiences and tools to be systematically provided? What systems of evaluation can adequately judge student progress in achieving goals toward which he is working? How can the evaluation systems be effectually geared into student counseling arrangements and admissions processes? How can program revision be made on the basis of evaluation, if indicated?

All these are key questions that require solutions if a university is to be viewed and operated as a unified learning environment. This is one way of thinking about a university operationally that might help develop new ideas on future growth patterns. I have, for the most part, avoided in this simple essay any attempt to explore the graduate education, research, and public service functions of institutions.

Now let me set forth a series of propositions which provide one

possible university model at the undergraduate level in answer to most of these questions. These propositions fit into an operational system. In a sense they provide one way of defining organizational content and policy for a university model.

1. The highest performance of the educational functions of a university depends on finding the most effective allocation of personnel, physical and financial resources.

2. The prime ingredient of university excellence is the capacity of its faculty to perform at the top levels of proficiency in the functions undertaken by the university.

3. The university environment must provide the tools, the time and the incentive for the continuous growth of faculty excellence and its demonstrated capacity to perform increasingly higher levels of accomplishment.

4. The most effective organization and conduct of the learning environment can be achieved by faculty effort to design activities which progressively require the student to assume more responsibility for learning by independent study and growth.

5. The organization and conduct of a program of independent study requires that the institution become centrally concerned with the application of principles of learning to the mastery of general and specialized areas of knowledge.

6. A university environment to facilitate independent learning requires that the university assume a major responsibility for assisting students develop the *skills and motivation for learning how to learn with increasing independence.*

7. A major program of independent study would include provisions for the student to progress through his program of study at a speed geared to his own achievement rate, beyond a minimum level of accomplishment required for continuing on the rolls of the university.

8. At the heart of a university program of independent study will be a comprehensive system of guidance and counseling requiring the closest union of professional skills in these fields with the competence of the academic faculty.

9. College and departmental objectives developed by the faculty should be specific enough to allow the derivation of the major blocks of subject matter, experiences needed for their mastery, and the devising of a comprehensive system for measuring student progress and accomplishment.

10. The process of deriving subject-matter blocks, experience needed, and the system of evaluation should be oriented around the following definition of elements:

(a) What principles are to be mastered?
(b) What understandings are to be achieved?
(c) What attitudes and values are to be considered?
(d) What skills are to be acquired?
(e) What applications are to be demonstrated?

11. The definition of student experiences needed to master the elements of each course block should provide for the maximum use of learning aids, on the one hand, supplemented by the most appropriate use of faculty time and effort, on the other hand.

12. The types of professional assistance needed by students from the academic faculty and guidance and counseling personnel will require the careful definition of functions to be performed and individual and group situations in which the assistance can be provided.

13. The roles to be performed by academic personnel in conduct of the undergraduate program should exclude non-professional responsibilities and provide, to the extent practical, differentiation of functions among the academic ranks.

14. The university should develop a major installation or installations to provide a learning resources center as the facility for use of learning aids by students and the developing of learning materials by the faculty. This would include the library, television, films, teaching machines, film strips, and so on.

15. As a part of the learning resources center, there should be operated with necessary studio and recording equipment either an open circuit UHF installation or closed-circuit facilities, capable of simultaneous telecasting on no less than six channels for 16 hours per day devoted exclusively to instructional telecasts.

16. The major part of the lecture and demonstration work of each unit of the university should be conducted by television, to the extent appropriate, and recorded on video tape to provide flexibility in scheduling and to save professional time from repetitious lecturing and demonstration.

17. There should be an exploration of the feasibility of having the basic continuity of lectures and demonstrations for each large-block course telecast three days a week with the class contact hours, two hours per week, designed for discussion and amplification of the material covered in independent study.

18. The faculty in each unit of the university should have the continuing responsibility for program development, evaluation, modification of the learning aids for their program with time and reward being provided staff for this function.

19. The living environment and the organization of student

life should be designed to complement the independent study orientation of the university program.

20. The development of academic facilities and residential facilities for students should be designed to facilitate the various aspects of the university educational program.

21. The full impact of the program envisioned in the preceding propositions should be to provide a more effective academic program for the money spent as compared with conventional programs. It would *reduce:*

(a) the number of courses in the university curriculum
(b) the number of hours students spend in formal class contact with faculty
(c) the time required by students to get a degree, particularly for the average and exceptional student
(d) the time required of faculty for repetitious lecturing and demonstration
(e) the time of faculty spent on non-professional activities
(f) the cost of physical plant needed for instruction

22. The program would *increase* the following factors:

(a) the initial time faculty would devote to development of educational objectives, course blocks, learning aids, and methods of evaluating student progress
(b) the facilities and time required for student counseling and guidance
(c) the facilities and personnel needed to help students improve learning skills
(d) the university investment in the development of learning aids and their operation in the academic program
(e) the time of faculty for research and writing
(f) the number of student credit hours per faculty member employed
(g) the salary of faculty members

The undergraduate program that would emerge from the assumptions set forth above would make possible an increase in the depth and variety of educational experience for each student if the goal of the university is to focus only on those learnings that are most relevant to developing the maximum intellectual power of the student in his general and specialized studies.

The efficient use of learning aids and professorial time should expedite student mastery of requirements and allow more time for exploration of specialties in depth as well as for studying a wider variety of subjects than is usually possible under conventional procedure. The dividing line between the method and pur-

pose of undergraduate education and any graduate programs that might be ultimately developed would be less sharply drawn in such a program.

The propositions set forth here are merely an exercise in university model making. You may not agree with the assumptions underlying any or all of them. There are many elements with which they do not deal. The financial implications or curriculum content are not detailed, for example. This much can be said, however: Given an institution with an enrollment of twenty to thirty thousand and a fixed per-student income, the patterns suggested here could double faculty salaries and provide for financing developmental work and for the operation of technological aids to learning.

The exploration of models such as this, and many others that could be devised, will be viewed by many as a fruitless enterprise. They will contend that to attempt a radical transformation of existing institutions is not realistic. This may be true now, but it need not be.

In addition to development of institutional models, equal effort should be given to the design of an evolutionary pattern for educational and financial change. Such patterns should be at least a ten-year projection of a growing educational organism. The university should have its projected growth pattern reflected in year-to-year budget planning with resource allocations consistent with the ten-year projections. Equally important is the development of a strategy and system for institutional change. A systematic plan of institutional change is missing in most higher institutions. On the contrary, in most institutions, the administration and faculty have developed and perpetrated a relationship which renders impossible modifications of any great significance.

Faculty, administration, and trustees should explore different ways to negotiate the means for major moves in institutional educational growth. Traditional modes of institutional government will probably not suffice.

The validity of the ideas we have explored here need to be tested further. One prospect seems relatively assured. Either the colleges and universities are going to change themselves according to competent professional designs, or their future form and operation will be decided at ever-increasing rates by public pressure and legislative action.

ROBERT L. THORNDIKE,
Teachers College,
Columbia University

In Memoriam: Irving D. Lorge

WE MEET this year at the usual time and place for the Invitational Conference on Testing Problems. However, a familiar face is missing today for the first time since these annual meetings have been held. Dr. Irving Lorge, my Teachers College colleague, whom many of you heard on the program a year ago, died suddenly last January of a heart attack.

Dr. Lorge had been associated with the Educational Testing Service and the College Entrance Examination Board in many different capacities over a long span of years. He appeared as a speaker at these meetings in 1948, 1952, 1953, and 1960, and he was chairman of the 1956 Conference. He was a member of the Educational Testing Service Standing Committee on Research from its inception in 1948 to the time of his death. He was also a member of the Advisory Committee for the Academic Ability Test and chairman of its Subcommittee on Verbal Comprehension.

For the College Entrance Examination Board, Dr. Lorge had served as a member of the Committee on Research and Development since 1955 and as chairman of this Committee since 1960. He was also a founding member of the Committee of Examiners on Aptitude Testing.

It is impossible to encompass Dr. Lorge's far-ranging activities and interests in a few words. He has worked on problems of learning theory, the description and education of the gifted, readability, vocabulary and semantic counts, group problem solving, statistics and test theory, machine computation, and even sugar chemistry. He was interested in research problems, whoever was undertaking them and wherever they were found, and he was interested in the students or colleagues who were working on these problems. He was a rigorous critic, but a strong supporter in time of trouble.

We shall all miss him as an intellectual leader and a personal friend.

Theme:
Extended Conceptions
of Evaluation
in Higher Education

Measuring the Quality of a College or University

DEWEY B. STUIT,
College of Liberal Arts,
State University of Iowa

THE PROBLEMS encountered in measuring the attributes or qualities of an educational institution do not differ basically from those encountered in the measurement of the attributes or behavior of the individual. Members of the group here assembled have spent a great deal of time and effort in the development of instruments designed to measure behavior and in the interpretation of the results obtained by use of these instruments. In assessing the quality of a college or university, as in measuring the behavior of an individual, one must first decide upon the dimensions to be measured, then the necessary instruments must be developed and, finally, appropriate reference points or bench marks must be established for the interpretation of results. In a very real sense, of course, the behavior of an institution is ultimately determined by the behavior of the individuals associated with that institution. For these several reasons, therefore, the measurement of the quality of an educational institution does not confront one with tasks which are totally new and different and much of what we have learned about measurement in recent decades has direct application to the assessment of the quality of our colleges and universities.

A major thesis of this paper is that the quality of an institution should be determined by measuring its products—what the institution does by way of developing the talents and attributes of its students and what it does to encourage productivity in its faculty. The primary business of colleges is the education of students; in the case of universities, equal emphasis is usually placed upon

The writer wishes to acknowledge his indebtedness to Dr. Norman Frederiksen and Dr. Jerry Helmstadter who collaborated with him in the writing of the report, *Survey of College Evaluation Methods and Needs*, submitted to the Carnegie Foundation. The present paper is based in considerable part on that report.

scholarly productivity and perhaps to a somewhat lesser degree, special services rendered to the state and nation. Whatever the products of an institution may be, their quality (and sometimes quantity) should be measured to determine the institution's adequacy or excellence. It isn't the beauty of the ivy on the walls, the splendor of the buildings, or even the size of the budget that counts; rather, it's what happens to students, and in the case of universities, what they produce by way of research and the extra services they render to society.

Having said this, the next question is how does one measure these products? The answer to the latter question is far from simple. First, one must know the objectives of the institution. What changes in student behavior does the college or university aim to produce? What level of competence are the students expected to reach? What emphasis does the institution place upon scholarly productivity? What services does it aim to provide for the community—city, state, region, or nation?

Second, one must reach a decision as to the measuring instruments to be employed. Will the chief reliance be placed on tests which purport to measure changes in student behavior—in knowledge, critical thinking, attitudes, and appreciations? Will the "success" of students following college graduation be used as a measure of quality? Will research productivity be measured by the number and quality of books and articles written by the faculty? Will service be measured by the number of speeches given, conferences attended, or letters written? Or, instead of using these more or less objective indices, will outcomes be measured or assessed exclusively by the intuitive judgments of those deemed to be competent judges?

Third, one must allow for the quality of student input. A college or university with a highly selective admissions policy will show up with better results in achievement tests and will graduate more illustrious alumni, not necessarily because of the high quality of education at that institution but at least in part because of the high level of talent enrolled. Very successful coaches of intercollegiate athletics not infrequently admit that the "secret" of their success is recruiting. Likewise, one of the major reasons explaining the success of a college's alumni is careful selection of the freshman class. This problem, namely, the differences in quality of input, can of course, be dealt with statistically; nevertheless it is a factor which must be recognized in institutional evaluation and has been seriously neglected in some earlier studies.

One could proceed indefinitely to recite the many problems

involved in trying to measure objectively the products of institutions. These problems are formidable and one could rather easily convince oneself that the difficulties inherent in this approach are insurmountable. Many of our institutions of higher education, for example, are liberal arts colleges, hence one ought to have available some means of measuring the major outcomes of a liberal education. Likewise, there is need for suitable instruments to measure instructional outcomes in each of the many professional programs found on the campuses of our colleges and universities. Admittedly these measurement problems are difficult and complex, but perhaps no more complex than the problems which exist in other fields. Placing a satellite in orbit would probably have been looked upon as a well-nigh impossible task twenty-five years ago. Now we take it as a matter of course. In my judgment there is no overriding reason to believe that the measurement problems involved in the evaluation of colleges and universities are so difficult that they cannot be solved in due time.

While I would hope that our long-range goal would be that of developing the necessary instruments and techniques for assessing the quality of an institution by measuring directly the quality of its products, at the moment we do not have the necessary instruments and techniques which would permit us to use this approach exclusively. I believe that at the moment we could use that method in part, but we must also use other approaches.

If one cannot do an acceptable job of assessing the quality of an institution by measuring the quality of its products directly, the next best thing is to measure the factors or dimensions which appear to correlate significantly with the quality of the product and which are educationally meaningful. Common sense, plus supporting evidence from such studies as reported in *The Evaluation of Higher Institutions*,[1] suggests first that certain factors or dimensions appear to be rather closely associated with institutional quality and are educationally meaningful. Among these factors are: (1) the objectives, (2) curriculum, (3) experience and training of the faculty, (4) quality of teaching, (5) aptitude and achievement of students, (6) number and quality of personnel services, and (7) adequacy of the library. There is also some evidence that amount and quality of institutional research (reference is not being made here to research in subject-matter fields) and institutional tone or atmosphere are related to academic quality. In addition, such factors as financial circumstances, physical plant, administrative organization, and athletic policy have been

studied as dimensions of institutional quality, particularly by accrediting agencies.

If one were interested in the maximum economy of effort in evaluating an institution, it would seem best to concentrate on measuring the quality of the faculty and the changes which occur in students. However, since the nature of what one seeks to measure in the faculty and students is conditioned by the institution's objectives, one can hardly omit the statement of objectives from the evaluation process. Because of their positive relationship to the quality of an institution, it seems appropriate to spend a few moments discussing the three dimensions mentioned, plus one other—intellectual atmosphere.

Objectives. The curriculum, the student personnel services, library resources, laboratory equipment, and extracurricular activities—in short, the entire educational program of a college or university should be an outgrowth of its educational objectives. Although nearly every college catalogue contains some statement of objectives, it is not always clear that the curricular offerings and related educational activities are planned to implement these stated objectives.

It would seem, therefore, that one of the first steps in any program of institutional evaluation, whether conducted by the institution itself or some outside agency, should be that of assessing the quality of its stated objectives. Among the critical questions which should be asked are the following: (1) Are the institution's objectives appropriate for the student body and community which it seeks to serve? (2) Are its objectives stated with sufficient precision so one can determine what student behaviors the institution is seeking to develop? (3) Were sound procedures used in arriving at its objectives? (4) Are the institution's objectives expressed clearly enough so they can be translated into curricular offerings and other types of learning experiences? (5) To what extent is the institution, particularly its faculty, concerned about its objectives?

Any institution of quality should have a statement of objectives which can be examined or inspected for adequacy—but this is not enough for purposes of evaluation. What institution "A" needs to know is how appropriate its objectives appear to be when compared with those of institution "B"; this holds true for each of the other dimensions by which its objectives can be measured. Although it is sometimes said that each institution should be evaluated with respect to its own objectives, no institution is in a class completely by itself. It can always be compared with one

or more quite similar institutions. Colleges and universities should be evaluated in terms of the objectives which they are seeking to achieve, but this should be accomplished by grouping institutions with similar programs and purposes and then making comparisons within these homogeneous groups. In my judgment it is incorrect to say that an institution is so unique that it cannot be compared with any other institution.

Faculty. If the most distinguishing characteristic of an institution is its curriculum, it could be said, perhaps, that its greatest claim to fame is the quality of its faculty.

The chief questions which should be asked in evaluating a particular institution are the following: (1) How do the training and experience of its faculty compare with those of sister institutions? (2) How do its faculty members compare in scholarly productivity with those in similar institutions? (3) How do the attitudes of its faculty members toward scholarship, teaching, and students compare with those of other faculties? (4) How is the morale of its faculty? (5) How do the community, professional, and other service activities of its faculty compare with faculty members elsewhere?

The individual administrator or faculty committee wrestling with the problem of faculty evaluation works under a considerable handicap. The experienced administrator is able to draw upon his own observations of colleges, but in most cases such observations are limited to a small and possibly unrepresentative list of institutions. For this reason normative data collected from sister institutions can be very helpful in making a self-study of the faculty and also in the appointment and promotion process.

A praiseworthy beginning was made by the North Central Association in the 1930's in providing data of this sort, particularly with reference to such factors as number of Ph.D.'s on the faculty, years of teaching experience, and number of scholarly papers and books published. More recently completed tabulations have appeared from time to time in the *North Central Association Quarterly*. While simple, tabular data regarding academic degrees, years of experience, and number of scholarly publications produced by the faculty may be subject to misuse, the availability of such statistics would be generally helpful. At the present time an individual institution can make such comparisons only by examining college catalogues or by writing letters to many individual institutions. If these statistical data could be supplemented with some attitude and morale factors regarding faculty members, a

genuinely helpful set of data would be provided for purposes of evaluating colleges and universities.

There seems little doubt, however, that we need better methods of evaluating faculty and administrative personnel than are generally available at the present time. The work of Clark[2] in identifying the members of the American Psychological Association who were judged to have made the greatest contribution to their profession, would seem to offer considerable promise. This type of study ought to be carried out in a number of other areas. The results of a comprehensive study of several professional fields would be very helpful in assessing the quality of faculties, especially at our universities.

Students. It is a well known fact that the quality of students as measured by scholastic aptitude tests varies considerably from college to college. Although most colleges and universities would prefer to enroll students of high scholastic aptitude, it is generally agreed that higher education should not be limited exclusively to the intellectually elite. One of the chief characteristics of American higher education is its diversity—in purposes, curriculum offerings, physical facilities, and also in types of students enrolled—but this does create problems in evaluating institutions.

Some of the questions which may well be asked of an institution regarding its students follow: (1) Considering its location, facilities, and purposes, is it drawing the size of student body which would be expected for an institution of its type? (2) How do its students compare in academic aptitude with the students enrolled in institutions in its classification? (3) Assuming that its courses and teaching procedures are the equivalent of those in sister institutions, how do the achievements of its students compare with those in other schools—test scores, honors won, number going on to graduate and professional schools, etc.? (4) How is the morale of the students in comparison with students on other campuses? (5) How do the educational, social, and cultural backgrounds of the students compare with those in sister institutions?

In the early years of the use of the American Council on Education Psychological Examination, reports of the average scores at different institutions were published in the *Educational Record.* After a few years this practice was discontinued and at the present time the general policy is not to publish the average scores of students at specific institutions. In some state and national testing programs it is, of course, quite easy for the individual college or university to determine how it compares with other institutions of its type, but this information is not usually available to the

general public. Granted that it might not be advisable to publish, by name, the average scores in aptitude and achievement tests for students at different institutions, it would be extremely helpful to have such information available in a handbook on an anonymous basis—grouped by types of institutions. The same would be true regarding drop-out rates, honors won by students, success of graduates, etc. The availability of such data would be very helpful in institutional evaluation.

Intellectual Atmosphere. Over and above the quality of the faculty, the richness of the curriculum, the adequacy of the physical facilities, and possibly even the effectiveness of teaching in the classroom, is a factor or dimension which one may call the intellectual climate prevailing on a particular campus. Although this dimension may be somewhat vague or nebulous, experienced educators would probably agree that there are noticeable differences in the intellectual tone of our campuses. The recent work of Pace and Stern[3] and Thistlethwaite[4], among others, lends support to the view that the climate which prevails on the campus is related to the quality of the institution's work, at least as measured by such a criterion as Ph.D. productivity.

Illustrative of the sorts of questions which pertain to the intellectual climate of a campus are the following: (1) Do students, faculty and administration appear to place intellectual attainment at the top of their scale of values? (2) Are extracurricular and service activities given priority over academic attainments? (3) Do students and faculty feel that they have freedom to express themselves without fear of censorship? (4) Is the unorthodox point of view welcomed on the campus or is it only tolerated or even discouraged? (5) Is there much campus debate or discussion about serious intellectual matters or fundamental issues facing our society?

Up to the present time no association of colleges or universities or any accrediting agency has made a serious attempt to assess this nebulous quality of a total institution. It is, of course, the sort of dimension which no one institution can very well study by itself. Any single college or university making a serious self-study may gain some insight into the intellectual tone of the campus, but it would be very easy to miss or overlook this quality unless one compared the situation prevailing on the local campus with that existing elsewhere.

Since it is generally recognized that the intellectual atmosphere varies from campus to campus and that this factor may be important in its effects upon the work of the institution, it would

seem that more systematic efforts should be made to collect information about it and to publish it in such a way that each college or university could tell about where it stood with respect to this dimension. Samplings of faculty and student opinions obtained by some of the instruments recently developed, reports on the sorts of administrative controls which are exercised, statements of factors considered in the appointment and promotion of faculty members, and similar types of data could be collected and evaluated so as to place a particular institution on a scale of "goodness" or "badness" of the intellectual climate prevailing on the campus.

In recent years colleges and universities have been questioned as to the degree to which they have encouraged serious intellectual work. We ought to have data which provide answers to such questions. Fortunately, some excellent research on this subject has been completed during the past five or six years and is going forward at the present time.

The principal techniques which have been employed to date in institutional evaluation are the personal interview and the questionnaire. Accrediting agencies, for example, have required institutions to supply an extensive body of information by answering questions put to them either in oral or written form. Judgments concerning the quality of institutions have then been reached in large part on the basis of the information collected through personal visits and the use of questionnaires.

Despite the fact that aptitude tests, achievement tests, attitude scales, and similar types of measuring instruments have been widely used in higher education, the results obtained by means of these instruments have not been used systematically in institutional evaluation. There is considerable evidence in the literature of personnel work which casts doubt on the adequacy of the interview and related observational procedures as data-gathering devices—and yet, these methods more than any others, have been, and are being used in institutional evaluation, particularly in accrediting. Inevitably, some use of interviews and observational procedures, and extensive use of questionnaires will always be required. In addition, if colleges and universities are to be given assistance in self-evaluation, they should have available the results obtained from large-scale testing programs so the individual institution can see how it compares with sister institutions. The individual or collective judgments of one or more interviewers are simply not adequate for this purpose.

In recent years a number of follow-up studies have been made

which would seem to have a considerable application in institutional evaluation. Particularly to be noted are the studies by Knapp and Greenbaum,[5] Knapp and Goodrich,[6] and Havemann and West.[7] Holland[8] and Thistlethwaite[4] have made some significant studies which take into consideration the role of student aptitude in the evaluation of institutions.

Many individual colleges and universities have made significant follow-up studies of their own graduates. Very frequently, however, these studies have been incomplete and have failed to serve their purpose because of lack of normative data. If, for example, a particular liberal arts college discovers that thirty per cent of its freshmen do not return for the sophomore year, that fifteen per cent of its graduates go on for graduate work, that twenty-five per cent of its applicants for medicine are accepted for admission, and that ten per cent of its graduates go into teaching, what do these statistics mean? The answer is that they are not very meaningful until compared with the performance of students at similar institutions. When studied in the light of "norms of performance," they become meaningful and useful not only to the institution itself, but also to prospective students.

Obviously, there are many problems involved in making over-all evaluations of institutions. One of the problems has been referred to earlier, namely, the different levels of student scholastic aptitude in different institutions. This problem can be dealt with statistically as shown by Holland[8] and Thistlethwaite.[4] Another problem is that of differences in institutional objectives. Again, this can be dealt with by a very careful homogeneous grouping of institutions having similar purposes and programs. Still another is the lapse of time factor. However, if it is agreed that institutions should be evaluated on the basis of their products, that is, the extent to which the "aimed for" changes in student behavior are achieved, then it would seem that evaluations of the follow-up type should play an important role in measuring the quality of colleges and universities.

The four dimensions of institutional quality which have been briefly discussed in this paper do not exhaust the list of factors which are positively correlated with the quality of the institution's products. Adequate evidence concerning these four factors, particularly the quality of the faculty and the performance of students, would go far toward making possible valid and reliable evaluations of colleges and universities. The major problems are those of devising adequate measuring instruments and of assembling the necessary normative data.

page 83

As stated earlier in this paper, the interview and the questionnaire have received the most extensive use in institutional evaluation and these techniques will continue to be used in the future. A great deal of research has already been done on interviewing, but not in relation to institutional evaluation. Questionnaires have also been studied extensively, particularly with reference to the wording of questions and the recording of data. No doubt a good deal could be done to improve the questionnaires used in evaluation studies, checking the validity of items, weeding out the nonessential, and facilitating the recording of data.

In the North Central study of the 1930's, *The Evaluation of Higher Institutions*,[1] a good deal of use was made of ratings. There would seem to be a continuing place for ratings, particularly when evaluating the institution's objectives, curriculum, and personnel services. The type of rating device to be used, the instructions prepared for evaluators, and the conversion of ratings to some scale score such as percentiles are all problems which are deserving of study. Since rating scales call for human beings to serve as measuring instruments, it is important that more research be done on ways and means of preparing the raters to perform their particular rating tasks.

In addition to improved methods of observational assessment there is an urgent need for instruments which will measure the changes in behavior which occur in students as a result of their educational experiences. Some tests of this nature are already available but much could be done to improve them for use in institutional studies. In the case of liberal arts colleges, for example, it would be highly desirable to have available a test battery measuring the major outcomes of a liberal education. Intensive work ought to be started on such a test battery; similar instruments are needed in many professional training programs.

Development and experimentation should not be limited to tests which measure intellectual outcomes alone, although these tests should receive the primary emphasis. There should also be better instruments to measure changes in attitudes and appreciations occurring during college years, for these are the outcomes frequently stressed in liberal education. These types of instruments may not yet be ready for wide-scale use, but intensive research on them should be encouraged.

Finally, there is a great need for normative or comparative data. Some of these data consist of nothing more than enumerations or listings of various items—withdrawal rates of students, degrees held by faculty members, number of students going on to

graduate school. Test scores constitute another form of normative data. The important point is that the collection of these data, arranging them in appropriate form, and then distributing them to the members of homogeneous groups of institutions would serve a very useful purpose. Quite apart from possible uses in institutional evaluation such as accrediting, the availability of such normative data would be extremely helpful to institutions in making self-studies.

I now return to the note on which I started, namely, that the measurement problems involved in the assessment of individuals and of institutions are not basically different. In both instances we need definitions or working hypotheses as to what we are seeking to measure, we need effective measuring instruments, we need normative data, and we need interpretative skill in arriving at a meaningful interpretation of the results. Why, then, may we ask, despite the excellent research completed in recent years, haven't we made more progress in institutional evaluation? Perhaps the answer lies in certain peculiarities which distinguish institutional from individual evaluation.

First, within our own circle, perhaps it should be admitted that we in higher education have evidenced a certain unwillingness to face facts as objectively as we should. Most of us like to feel that we are doing a good job and, at times, conveniently overlook the defects which exist in our own institutions. If we were to gather the sorts of data suggested in this paper we might disclose some facts which would not be welcomed by many American colleges and universities.

Second, I think we have a fear, partly legitimate, that data of the sort described in this paper would be subject to misuse. A failure to interpret the data properly, or a twist of emphasis here or there, intentional or unintentional, might result in harm to a particular institution. Actually, I believe that this danger should be accepted as a challenge on our part to do a better job of explaining higher education to the public.

Third, there is the problem of our measuring instruments. Our present procedures for the assessment of institutional quality are lacking in precision. A good deal of research should be done on them. Nevertheless, one should not wait for the development of perfect instruments. If this were done, we would never make progress in institutional evaluation.

Finally, there is a reluctance on the part of institutions, including some very good ones, to devote time and effort to collect the sorts of data needed in a thorough evaluation. Even though it

might take some class time and some extra consultation and deliberation, I believe that time spent in careful evaluation would be well spent.

The problem before us is a challenging one. We in America have some of the best colleges and universities in the world. We also have our share of poor ones. Our country has great need for quality in higher education. I believe that institutional evaluation such as proposed in this paper would help to raise the level of our colleges and universities and enable them to contribute more significantly to our strength as a nation and to our way of life.

REFERENCES

1. *The Evaluation of Higher Institutions*, Volumes I, II, III, IV, V. Chicago: University of Chicago Press, 1937.
2. Clark, Kenneth E. *America's Psychologists; A survey of a growing profession*. Washington, D. C.: American Psychological Association, 1957.
3. Pace, C. Robert and Stern, George C., "An Approach to the Measurement of Psychological Characteristics of College Environments." *Journal of Educational Psychology*, 49, 269-77, October, 1958.
4. Thistlethwaite, Donald L., "College Environments and the Development of Talent." *Science* 130:71-76, July 10, 1959.
5. Knapp, R. H. and Greenbaum, J. J., *The Younger American Scholar: His Collegiate Origins*. Chicago: University of Chicago Press, 1953.
6. Knapp, R. H. and Goodrich, H. B., *Origins of American Scientists*. Chicago: University of Chicago Press, 1952.
7. Havemann, Ernest and West, Patricia S. *They Went to College*. New York: Harcourt, Brace and Co., 1952.
8. Holland, John L., "Undergraduate Origins of American Scientists," *Science* 126:433-37, September, 1957.

The Nature and Use of Institutional Research

JAMES I. DOI,
Institutional Research,
University of Colorado

WHAT IS institutional research? Is it simply a fancy, all-embracing name for research in higher education, or is it a particular type of research concerned with special problems of colleges and universities? Is it a new type of research activity or is it merely a new name for something which colleges and universities have been doing for many years? Is it applied research, basic research, or both? Is it another short-lived educational fad, or can it become a significant feature of American higher education?

Five years ago when I joined the University of Colorado as director of its newly established office of institutional research, these and similar questions seemed important to me, and I sought to formulate some answers. I found this task both disappointing and encouraging. I was disappointed to find so few published statements on the nature and role of institutional research. Ruth Eckert and Robert Keller's collection of University of Minnesota studies,[1] published in 1954, was at that time the only major work recently published. There were also a number of institutional surveys, which one might construe to be a form of institutional research although the authors themselves did not refer to them as such. Then there were the many collections of fiscal and enrollment data, analyses of student characteristics, and studies of instructional programs. These reports and studies emanated from the business office, the registrar's office, the office of the dean of students, and the instructional departments. Might not these also be construed as forms of institutional research? Or should institutional research be defined as something other than the kinds of data collections and analyses produced by these various university offices?

The paucity of published statements on the nature and role of institutional research was, at the same time, encouraging. I found that in suggesting policies for the development of institutional

research at the University of Colorado I could entertain almost any reasonable notion of what the program should be like without fear of authoritative challenge. The faculty committee which had recommended the establishment of the office of institutional research was generous in suggesting its functions. The committee had suggested that the director of institutional research: "(1) conduct research related to legislative appropriations; (2) conduct research related to any phase of university education, administration and operation: (a) in exercising this function, the director shall secure qualified personnel either from the university staff or outside sources, (b) shall make continuing research studies relative to effective control and general university improvement; and (3) conduct any other type of research required to meet the needs of the university."

One could not ask for a broader license. Any aspect of the university was fair game for the director of institutional research. But this represented only the view of one university.

In 1957 I doubt that there were more than six universities and a handful of liberal arts colleges and state colleges with formally organized offices or bureaus of institutional research. Since then, such units have been created in a dozen or more universities, among them Michigan State, Wisconsin, Wayne State, Utah, Arizona State, Tennessee, California, New York University, and Rhode Island. Such offices have also been recently established in liberal arts colleges and state colleges, but as to their approximate number, I have no idea.

Also in the last three or four years, several definitive and authoritative statements on institutional research have become available. In 1959 Professor W. H. Cowley of Stanford University gave his lecture entitled "Two and a Half Centuries of Institutional Research."[2] That same year, John Dale Russell published his ideas on the purpose and organization of institutional research.[3] In 1960, A. J. Brumbaugh expressed his thoughts on the subject in the pamphlet *Research Designed to Improve Institutions of Higher Learning*.[4] In 1959, W. Hugh Stickler of Florida State University completed his survey report of institutional research concerning land-grant institutions and state universities,[5] and the following year the Western Interstate Commission for Higher Education published Hall Sprague's study of institutional research in the West.[6]

Also, in the summer of 1959, the Western Interstate Commission for Higher Education launched its series of annual workshops on institutional research for college and university administrators

in the western states; the Southern Regional Education Board started its series of workshops for the southern states in the summer of 1960; Paul Dressel and his colleagues at Michigan State University last fall sponsored a workshop on institutional research for liberal arts colleges; and this summer, B. Lamar Johnson and his group at the University of California at Los Angeles organized a workshop on institutional research for junior colleges. Earlier this month the American Council on Education devoted a section of its annual meeting to a discussion of institutional research—its functions, organization, and future outlook.

What we are witnessing is the spring of institutional research. And as in The Song of Solomon: "...the time of the singing of birds is come, and the voice of the turtle is heard in our land." Thus, if I seem occasionally to chirp, I trust that you will be tolerant of my enthusiasm. And with apologies to W. H. Cowley, John Dale Russell, A. J. Brumbaugh, and Hall Sprague, let me review briefly what the turtle has to say.

W. H. Cowley defines institutional research as follows: "Educational research...consists in any thorough-going investigation of an educational problem; and institutional research in higher education constitutes such an investigation of any topic concerning which a college or university, or a group of them, collects or seeks to collect data toward the end of improving operations."[2]

This definition has three significant aspects. One, "any topic" can be a subject of institutional research. Two, institutional research can be undertaken by a single institution or collectively by a group of institutions on a topic of mutual interest. And three, the object of institutional research is to improve operations.

In his paper, Professor Cowley deplores the rather restricted current definitions of research, namely that research to be considered research must employ statistical concepts. He suggests that many educational questions of major gravity are not subject to measurement. By this, I gather, he is telling us that institutional research in higher education can and should be concerned with non-statistical studies, as well as those involving statistics. With this I agree.

It is also interesting to note that Professor Cowley had been asked to review in his lecture the three most recent decades of institutional research in American colleges and universities, but he chose to speak on two and a half centuries of institutional research. It was his way, I suspect, of telling those of us who were new to the game and therefore prone to think of our efforts as being new, that from the very beginning American colleges and

universities have been engaged in institutional research. Some of the terms and methods used may be relatively new, but the nature of the activity is very old.

John Dale Russell[3] views institutional research as a form of applied research. He notes that basic research is carried on solely in the spirit of inquiry, without regard to any practical use or application of the results. Institutional research, however, is nearly always stimulated by the desire to find an answer to some practical problem. Some basic research on higher education can arise out of institutional research, such as when the search for an answer to a practical problem requires supplementary research on the techniques of investigation or even the invention of new techniques.

There is a tendency in institutions to assign to the office or bureau of institutional research the responsibility of answering questionnaires and preparing statistical reports. Dr. Russell, in speaking about this point, says that this is not "research," if it involves only the compilation of readily available data. Thus, not every task assigned to an office of institutional research is research.

A. J. Brumbaugh advocates a somewhat broader definition of institutional research than does Dr. Russell. Dr. Brumbaugh gives this definition: "Institutional research...consists of studies and investigations focused on current problems and issues in institutions of higher education. It also consists of studies and investigations of problems and issues that are basic to long-range planning or that may ultimately have implications for institutional operations. The former is typically applied research; the latter is basic research in higher education."[4]

By way of indicating what he means by basic research in higher education, Dr. Brumbaugh refers to research in specific disciplines that may have educational applications such as studies in the psychology of the learning process.

I shall not take the time here to discuss whether some of the differences in the definitions given by Drs. Cowley, Russell and Brumbaugh are more apparent than real. The important thing is that all three conceive of institutional research as investigations which cover a wide range of topics and which are designed to improve institutions of higher education.

Let me now return to the question of how new institutional research is. As Dr. Cowley has pointed out, American colleges and universities have been engaged in institutional research for a very long time. But many of us currently interested in institutional research believe that there is something new about it.

This "something new" has been aptly stated by Hall Sprague: "The term *institutional research*, as a collective designation for various kinds of institutional self-studies, is fairly new. Even newer is the notion that institutional research is, or should be, a recognized function of a college or university, analogous to business management, public relations or student personnel services. More recent still is the idea that this function should be coordinated and/or reviewed by a person or office specifically designated for this purpose."[6]

It is the coordination and systematic review of research needed for institutional improvement and the recognition of this process as a continuing, day-to-day function of a college or university that gives contemporary institutional research a distinctive place in the history of educational research.

The days of the sporadic, one-shot institutional survey and the uncoordinated research efforts of the individual business manager, the registrar, the testing officer, and the dean of faculty are fast disappearing. The college or university of today is far too complex and is changing too fast to be adequately understood and managed by such efforts.

This does not mean that there will no longer be committee surveys or that the business manager, the registrar, the testing officer, and the dean of faculty, among others, will not be involved in institutional research. Indeed, they must and will be involved. But their efforts will be coordinated and the results of their research examined and synthesized so that the institution as a whole can move ahead on the basis of mutual understanding of research facts.

It has been frequently noted that a research-conscious, aggressive dean or business manager will move far ahead of the rest of the institution, sometimes so far ahead that the laggers impede further progress on his part. The resulting push and pull has been the source of much friction within institutions. Patience has its limits and sooner or later criticism of one's colleagues must flow.

Or, take another example. The testing officer and the registrar may have much data concerning student aptitudes, abilities, and personality characteristics. But these data have not always been brought to timely and effective use by faculty members in effecting changes in their instructional procedures and programs. The time lag between research completion and application of the findings may be as much as four or five years.

Coordinated institutional research operates on the assumption that not all administrative officers are equally interested in, or

capable of conducting studies of their operations. It has, as one of its objectives, the stimulation of operations research on the part of administrative officers. In an area where this attempt fails, it must find ways and means of getting the research done so that a failure in one particular area will not hinder general institutional improvement.

Coordinated institutional research also operates on the assumption that not all faculty members are equally receptive to self-examination and application of research findings to their instructional programs. Here again, it seeks to stimulate self-examination, to render research assistance where and when necessary, and to bring to the attention of faculty members institutional data pertinent to their functions. I suspect that if faculty members have at times seemed reluctant to engage in self-evaluation of their programs or to undertake educational experimentation, it is in part because the administration has not always provided them with the necessary assistance. Admittedly their reluctance also stems in part from the very human desire to protect sacred cows, but I believe we are in pretty general agreement these days that higher education cannot afford sacred cows fostered by self-interests.

Institutional research has already made good beginnings in clarifying many notions we have held concerning unit costs in higher education, space utilization, class size and teaching loads, student-faculty ratios, and scope of course offerings. These efforts were largely motivated by economy, and in a number of instances economy has been effected. But, institutional research is not concerned primarily with economy of operations. Its chief concern is institutional improvement, and this may, as it often does, mean increased financial support for many aspects of higher education.

The effective coordination of research designed to bring about institutional improvement is, I believe, good and sufficient reason for the current interest in institutional research. Two recent developments in higher education suggest other reasons for this interest. One is the increased concern on the part of colleges and universities for long-range planning. The other is that the current generation of social scientists has, in a sense, rediscovered the college student, the academic man, and college life as legitimate objects for research.

Let me first take up the relationship of long-range institutional planning to institutional research. I think many in higher education will concede that we have not planned our growth sysmatically and well. New programs have been initiated without adequate consideration to other more pressing needs of the insti-

tution or to long-term commitments and their implications. Certain institutions have permitted themselves to grow like Topsy, and while she may have grown to be an attractive woman because the laws of nature and the arrangement of her genes decreed that she should be so, there are no laws of institutional growth that guarantee an institution increased effectiveness along with increased size. If such effectiveness is to come, it must come through planning.

Some institutions have made a virtue of smallness—small enrollments, small classes, and small student-faculty ratios—and today find themselves near extinction. Even smallness must be planned in accordance with the facts of institutional resources and the expectations of students and faculty.

Most of the long-range planning that has been done so far has dealt with enrollment projections, finances, and physical plant. The usual procedure is to project a given enrollment, then compute an estimated budget for a given year by multiplying the number of students by an estimated per student cost. Physical plant needs are frequently computed by multiplying the estimated number of students by the number of square feet of floor space required per student. The more sophisticated planning studies are characterized by differentials in per student costs for graduate and undergraduate students, and differentials in space standards for graduate and undergraduate students and for various programs and functions.

The identification of these various differentials is the task of institutional research. But for one very critical aspect of planning—actually the most important aspect—institutional research has yet to come up with a satisfactory answer: How can an institution project changes in programs and functions? We all know that programs and functions change over a period of several decades, indeed sometimes change rapidly. A program which prospered a decade ago is less prosperous today, and another which was struggling for recognition has come to play a vital role in the life of the institution. A function which required only three per cent of the institutional budget two decades ago, today requires six per cent. Its space requirements have also doubled. Following are examples of programs and functions which have either declined or prospered in the last two decades—general education, international education, engineering, modern languages, sponsored research, the library, student services, graduate instruction and research, and intercollegiate athletics.

Yet you will find that many long-range studies of physical plant needs and financial needs are based on the assumption that in

1970 or 1980 institutional programs and functions will be pretty much the same as they are today. The planners know that this will not be so, but in the absence of knowledge of what the changes are likely to be, they have no choice but to proceed on outworn assumptions.

The search for this knowledge of what and when such institutional changes will occur and how they occur is, in my opinion, one of the most challenging tasks facing institutional research. Equally important is the search for knowledge on timing the changes properly and thus reducing the gap between what an institution is actually doing and what needs to be done.

A most heartening and promising recent development in higher education is the entry of a substantial number of psychologists, sociologists, and social anthropologists into the field of research on colleges and universities. This is particularly true in the areas of research on the college student, the faculty, and college life.

I do not mean to imply by this statement that social scientists have never before seriously regarded the college student, the faculty, and the campus as legitimate subjects for research. Social scientists of past generations were involved in research on higher education. Among them were Veblen, Robert C. Angell, Ruth Strang, and Willard Waller. And spanning the pre-World War II era and the current period were men such as Theodore Newcomb and Logan Wilson. But in general, previous generations of social scientists were preoccupied with the exotic and the unusual—the puberty rites of primitive groups, the impact of modern technology on an isolated mountain village, abnormal psychological behavior, causes of delinquency, the slums of a city, and the ethos of hobos. Only a few recognized that in their own backyard, the campus and classroom, were there appropriate subjects for research.

In perusing the recent literature on college students, I looked for statements which might help explain why the current generation of social scientists has taken the college student and college life as subjects for research. The Jacob's report[7] on the changing values of college students seems to have been an important stimulant to social scientists. His rather dismal conclusions concerning the impact of the educational process on student values jolted the academic community. Another stimulant is the availability of funds for research, especially in the way of foundation grants. Another is the development of computer techniques. Perhaps an even more important factor is the increased confidence that social scientists now have in their techniques of analysis and measure-

ment of social and psychological phenomena. This confidence is essential in analyzing situations and problems in one's own backyard, especially when that backyard is peopled by groups that are frequently suspicious, at times even hostile, and extremely vocal in their criticisms of both research methodology and results. The previous generation of social scientists was relatively immune from criticism—they took as their subjects primitive communities and deviant groups that cared little about research and were almost unknown to their colleagues. College students, however, do care about what the researcher has to say about what he thinks they think, and how they interact and feel; and faculty members also care very much about what a researcher may have to say of their students and themselves. Another reason seems to be that research on educational problems and issues did not in the past bring to the social scientist as much professional status as research on other subjects. Perhaps educational research carried with it the taint of being applied, rather than basic research.

Whatever the reasons, it is sufficient to note that many social scientists today are engaged in research in higher education and we can look forward to the results of their efforts.

Their works, however, pose a problem to the college administrator. The research findings may be difficult for him to absorb, not so much because of volume, which will be substantial, but because they will be written in technical jargon. Social scientists are no different from other scientists; they too tend to address their research reports to each other and not to the administrator or the laymen. Also the college administrator will find that a substantial portion of the social scientists' research findings fail in translation into significance for institutional planning and day-to-day operation. For example, what should and can an institution do about the fact that a high proportion of its students present a certain personality profile? Should it attempt to alter the composition? What purpose will be served by such an alteration? And how should it go about changing student composition if it should decide that it is desirable to do so?

The social scientists engaged in research on college students and college life are not unaware of the practical implications of some of their research for higher education and the fact these implications cause some anxiety among college administrators. For example, following is a statement which appeared in a recent publication dealing with the college student: "The question was raised this afternoon, what proportion of dissident or somewhat different students do you need in order to make their group strong

enough to sustain itself in a college, but not so strong that it tends to dominate the whole campus? This is a fascinating problem. I suppose it might be called human engineering, or something of the sort. It is somewhat frightening to think of setting up a human community according to some plan of that kind."[8]

This awareness of the implications of their research for the development of higher education is also manifested in the social scientists' willingness to participate in the recent regional workshops on institutional research. Many of the key speakers at these workshops have been social scientists. Among them were Theodore Newcomb and W. McKeachie of Michigan; Robert Pace and George Stern of Syracuse; and T. R. McConnell, Paul Heist, Burton Clark, and Martin Trow of the Center for Study of Higher Education in Berkeley.

Whether or not these social scientists use the term "institutional research" to describe their studies is unimportant. As a matter of fact, some of them do. What is important is that institutional research agencies consider it to be their task to work with these social scientists in interpreting the significance of their research to college administrators and in bringing to bear the new-found knowledge in improving institutions of higher education.

I have not in this paper attempted to describe specific topics of institutional research. This would be a fairly ambitious task, since institutional research, as previously defined, covers every study which is designed to improve a college or a university or which has implications for such improvement. The studies include investigations of unit costs, space utilization, faculty retention and withdrawal, student retention and withdrawal, faculty morale, student personality characteristics, the socio-economic backgrounds of students, the campus environment, instructional methods and procedures, program needs, administrative organization, relationship of faculty and administration, and so on.

I have chosen instead to emphasize the functional aspects of institutional research which are more important and pertinent to the subject than any collection of studies. Studies of colleges and universities will be made, as they have been in the past, regardless of whether one chooses to call them institutional research or some other name. As a function, however, institutional research seeks to bring about improvement in a college or university through continuous and systematic research effort.

As I pointed out earlier, what we are witnessing is the spring of institutional research. Whether it can effectively accomplish its tasks has yet to be fully known. In another four or five years, we

shall witness the summer of institutional research, and if its winter should come, I trust that it will only come because we have found a better method of effecting institutional improvement.

REFERENCES

1. Eckert, Ruth, and Keller, Robert (editors). *A University Looks at Its Program*. Minneapolis: University of Minnesota Press, 1954.
2. Cowley, W. H. "Two and a Half Centuries of Institutional Research," *College Self Study*, pp. 1-16. Boulder: Western Interstate Commission for Higher Education, 1960.
3. Russell, John Dale. "The Purpose and Organization of Institutional Research," *College Self Study*, pp. 17-22. Boulder: Western Interstate Commission for Higher Education, 1960.
4. Brumbaugh, A. J. *Research Designed to Improve Institutions of Higher Learning*. Washington, D. C.: American Council on Education, 1960.
5. Stickler, W. Hugh. *Institutional Research Concerning Land-Grant Institutions and State Universities*. Tallahassee: Office of Institutional Research and Service, Florida State University, 1959.
6. Sprague, Hall. *Institutional Research in the West*. Boulder: Western Interstate Commission for Higher Education, 1960.
7. Jacob, Philip E. *Changing Values in College*. New York: Harpers, 1957.
8. American Council on Education. *Spotlight on the College Student*, p. 51. Washington, D. C.: The Council, 1959.

PARTICIPANTS—1961 Invitational Conference on Testing Problems

ABRAHAM, Ansley A., Florida Agricultural and Mechanical University
AHMAD, Captain S. K., Indiana University
AHRENS, Dolores F., Educational Testing Service
ALBRIGHT, Frank S., West Orange (New Jersey) Public Schools
ALLEN, Margaret E., Portland (Maine) Public Schools
ALLISON, Roger B., Jr., University of Utah
ALMAN, John E., Boston University
ALT, Pauline M., Central Connecticut State College
ANASTASI, Anne, Fordham University
ANDERSON, G. Ernest, Jr., Newton Public Schools, Newtonville, Massachusetts
ANDERSON, Pauline K., New York State Employment Service
ANDERSON, Rose G., New York City
ANDERSON, Scarvia B., Educational Testing Service
ANDREWS, T. G., University of Maryland
ANGOFF, William H., Educational Testing Service
ARBOLINO, Jack N., College Entrance Examination Board
ARONOW, Miriam S., Bureau of Educational Research, New York City Schools
ARSENIAN, Seth, Springfield College, Springfield, Massachusetts
ATKINS, Joseph K., Educational Testing and Counseling, Baltimore, Maryland
ATKINS, William H., Yeshiva University
ATWATER, Stephen K., Knox College

BALL, Albert T., Monmouth College
BANNON, Charles J., Department of Education, Waterbury, Connecticut
BANNON, Marion K., Crosby High School, Waterbury, Connecticut
BARDACK, Herbert, New York State Department of Social Welfare
BARGMANN, Rolf E., International Business Machines Research
BARKER, Roland, Montclair (New Jersey) Public Schools
BARNETTE, W. L., Jr., University of Buffalo
BATES, Margaret A., Educational Testing Service
BENNETT, George K., The Psychological Corporation
BENNETT, Mrs. George K., Bronxville, New York
BENSON, Arthur L., Educational Testing Service
BERDIE, Ralph F., University of Minnesota
BERGER, Bernard, New York City Department of Personnel
BERGESEN, B. E., Jr., Personnel Press, Inc., Princeton, New Jersey
BERGSTEIN, Harry B., Simpson Junior High School, Huntington, L. I., New York
BERNIER, Joseph J., Pennsylvania State University
BIXLER, Harold H., Western Carolina College
BLANCHARD, D. D., Wilton (Connecticut) High School
BLEEKE, Donald E., Cranford (New Jersey) Public Schools
BLETTER, Robert P., Columbia University
BLIGH, Harold F., Harcourt, Brace and World
BOGART, Max, New Jersey State Department of Education
BOLLENBACHER, Joan, Cincinnati (Ohio) Public Schools
BORGATTA, Edgar, University of Wisconsin
BOYD, Joseph D., Illinois State Scholarship Commission
BRACA, Susan E., Oceanside (New York) Senior High School

page 98

BRANDT, Hyman, American Occupational Therapy Association, New York City
BREAM, Lois Gould, Cheltenham (Pennsylvania) High School
BRENNAN, Eileen F., Valley Stream North High School, Franklin Square, New York
BRETNALL, William B., Educational Testing Service
BRIDGES, Claude F., U. S. Army Enlisted Evaluation Center, Fort Harrison
BRODSKY, David J., Educational Testing Service
BROWN, Fred S., Great Neck (New York) Public Schools
BROWN, F. Martin, Fountain Valley School, Colorado Springs, Colorado
BROWN, W. R., University of Oklahoma
BRYAN, Glenn L., Office of Naval Research
BRYAN, J. Ned, Jr., United States Office of Education
BRYAN, Miriam M., Educational Testing Service
BUCHHEIMER, Arnold, The College of the City of New York
BUDDEKE, Sister Rita, Catholic University of America
BURKE, James M., Norwalk (Connecticut) Public Schools
BURKE, Paul J., Bell Telephone Laboratories
BURNHAM, Paul S., Yale University
BURNHAM, Philip, St. Paul's School, Concord, New Hampshire
BURR, William L., Harcourt, Brace and World

CAMPBELL, Donald W., Newark (New Jersey) Public Schools
CANNON, Deore J., The College of William and Mary
CAPPS, Marian P., Virginia State College
CARLSON, J. Spencer, University of Oregon
CARNEGIE, M. Elizabeth, Nursing Outlook, New York City
CARROLL, John B., Harvard University
CARSTATER, Eugene D., Bureau of Naval Personnel
CARSTATER, Marie H., Falls Church (Virginia) High School
CAYNE, Bernard S., Crowell-Collier Publishing Company
CHAPMAN, Gloria, Stanford University
CHAPPELL, Bartlett E. S., New York Military Academy
CHAUNCEY, Henry, Educational Testing Service
CHURCH, John G., New York State Education Department
CHURCHILL, Ruth D., Antioch College
CLARK, Philip I., California Test Bureau, Scarsdale, New York
CLEARY, J. Robert, Scarsdale (New York) Public Schools
CLEMANS, William V., National Board of Medical Examiners
CLENDENEN, Dorothy M., The Psychological Corporation
CLIFF, Norman, Educational Testing Service
CLIFFORD, Paul I., Atlanta University
COMBS, Charles, Board of Cooperative Educational Services
CONLAN, Gertrude, Educational Testing Service
CONNOLLY, John A., Educational Testing Service
CONRAD, Herbert S., United States Office of Education
COPELAND, Herman A., Pennsylvania State Civil Service Commission
CORNOG, William H., New Trier High School, Winnetka, Illinois
CORY, Charles H., Philadelphia Personnel Department
COWLES, John T., Maurice Falk Medical Fund, Pittsburgh, Pennsylvania
CRAVEN, Ethel Case, Polytechnic Institute of Brooklyn
CRAWFORD, John R., University of Maine
CREAGER, John A., National Academy of Sciences

CROOK, Robert B., Queens College
CROWE, Robert M., American College of Life Underwriters
CYNAMON, Manuel, Brooklyn College

DAILEY, John T., American Institute for Research
DART, Francis, National Academy of Sciences
DAVIDOFF, Melvin D., United States Civil Service Commission
DAVIDSON, Helen H., The College of the City of New York
DAVIS, Frederick B., Hunter College
DAVIS, Junius A., Educational Testing Service
DEAN, E. Douglas, Princeton, New Jersey
DEMAREE, Robert G., University of Illinois
DE MOOSE, Norman M., Archdiocesan Vocational Service, New York City
DERWIN, Edward P., Crosby High School, Waterbury, Connecticut
DESMOND, Helen E., Summit (New Jersey) High School
DETCHEN, Lily, Chatham College
DIAMOND, Lorraine K., Columbia University
DIEDERICH, Paul B., Educational Testing Service
DIGGS, Franklin B., New York City Department of Personnel
DION, Robert, California Test Bureau, Nashua, New Hampshire
DIZNEY, Henry, Kent State University
DOBBIN, John E., Educational Testing Service
DOI, James I., University of Colorado
DONLON, Thomas, Educational Testing Service
DOPPELT, Jerome E., The Psychological Corporation
DOWNES, Margaret, New York State Department of Civil Service
DRAGOSITZ, Anna, Educational Testing Service
DRESSEL, Paul L., Michigan State University
DU BOIS, Philip H., Washington University
DUGGAN, John M., College Entrance Examination Board
DUKER, Sam, Brooklyn College
DUNN, Frances E., Brown University
DUNNE, Reverend William J., S.J., National Catholic Educational Association
DUROST, Walter N., Test Center and Advisement Center, Dunbarton, New Hampshire
DUTTON, Eugene, University of Alabama
DVORAK, Beatrice J., United States Department of Labor
DYER, Henry S., Educational Testing Service

EBEL, Robert L., Educational Testing Service
EDGAR, J. W., Texas State Commission of Education
EDMONDS, W. S., Virginia State College
EDWARDS, Winifred E., Irvington (New Jersey) High School
EHRMANN, Irene E., Verona (New Jersey) High School
ELLIOTT, Godfrey, McGraw-Hill Book Company
ENGELHART, Max D., Chicago City Junior College
EPSTEIN, Bertram, The College of the City of New York
ERICKSON, Clifford G., Chicago City Junior College
ESSER, Barbara, Educational Testing Service
EVERETT, Margaret, Educational Testing Service

FAGIN, Barry, Marquette University
FAN, Chung T., International Business Machines

FARR, S. David, University of Buffalo
FELDT, Leonard S., State University of Iowa
FENDRICK, Paul, Western Electric Company
FENOLLOSA, George M., Houghton Mifflin Company
FENSTERMACHER, Guy M., Educational Testing Service
FIFER, Gordon, Hunter College
FINDLEY, Warren G., University of Georgia
FINEGAN, Owen T., Gannon College
FINK, David R., Jr., University of Maine
FISCHER, Clyde L., Puerto Rico Department of Education
FISHELL, Kenneth N., University of Rochester
FISHMAN, Joshua A., Yeshiva University
FITZGEREL, Robert K., Educational Testing Service
FLEISCH, Sylvia, Boston University
FLESCH, Mary H., Educational Testing Service
FLOYD, William A., Columbia College, Columbia, South Carolina
FORMICA, Louis A., Norwalk (Connecticut) High School
FORT, G. H., Atlanta (Georgia) Board of Education
FRANKFELDT, Eli, Human Factors Research Branch, United States Army
FREAS, Howard J., Jr., Hatboro-Horsham (Pennsylvania) High School
FREDERIKSEN, Norman, Educational Testing Service
FREEMAN, Paul M., Educational Testing Service
FRENCH, Benjamin, New York State Department of Civil Service
FRENCH, John W., Educational Testing Service
FRETWELL, Elbert K., Jr., New York State Department of Education
FRIEDENBERG, Edgar, Brooklyn College
FRIEDMAN, Jerry J., Valley Regional High School, Deep River, Connecticut
FRIEDMAN, Sidney, Bureau of Naval Personnel
FULTON, Ailsa W., American and Foreign Teachers' Agency

GANNON, Frederick B., Rochester (New York) City School District
GAINES, Lynn K., Educational Testing Service
GALLAGHER, Henrietta, Educational Testing Service
GARDNER, Eric F., Syracuse University
GEORGIA, Sister M., Rosary Hill College
GERBERICH, J. Raymond, University of Connecticut
GIBLETTE, John F., University of Maryland
GIDEON, Sara B., Pennsylvania Department of Public Instruction
GLASIER, Charles A., New York State Department of Education
GLASS, Gene, University of Nebraska
GODSHALK, Fred I., Educational Testing Service
GOLDMAN, Leo, Brooklyn College
GOLDSTEIN, Leo S., Center for Programed Instruction, Long Island City, New York
GOULD, Ben, Brooklyn College
GRAFF, Franklyn A., Westport (Connecticut) Public Schools
GRAY, Mrs. Lyle Blaine, Remedial Reading Consultant, Baltimore, Maryland
GREEN, Dorothy, United States Civil Service Commission
GREER, Harry H., Jr., United States Naval Examining Center
GULLIKSEN, Harold, Educational Testing Service
GUTHRIE, George M., The Pennsylvania State University
GUTHRIE, Helen, The Pennsylvania State University

HADLEY, Everett E., West Hartford (Connecticut) Public Schools
HAGIN, Rosa A., Irvington (New Jersey) Public Schools
HAGMAN, Elmer R., Greenwich (Connecticut) Public School
HALL, Alfred, Educational Testing Service
HALL, Robert G., Manter Hall School, Cambridge, Massachusetts
HALLENBECK, Edwin F., University of Rhode Island
HAMILTON, Mildred E., Queens College
HAROOTUNIAN, Berj, University of Delaware
HARTSHORNE, Nathaniel H., Educational Testing Service
HARTZELL, Karl D., Bucknell University
HARVEY, Philip R., Educational Testing Service
HAUCH, Charles C., United States Office of Education
HAYWARD, Priscilla, Educational Testing Service
HEATH, Robert W., Educational Testing Service
HEENAN, David K., Michigan State University
HEIDGERD, Lloyd, Educational Testing Service
HEIL, Louis, Brooklyn College
HELM, Carl E., Educational Testing Service
HEMPHILL, John K., Educational Testing Service
HERGE, Henry C., Rutgers, The State University
HERMAN, David O., The Psychological Corporation
HERMAN, M. L., Moses Brown School, Providence, Rhode Island
HERRICK, C. James, Hartwick College
HESLIN, Phyllis, National League for Nursing
HIERONYMUS, Albert N., State University of Iowa
HILLS, John R., University System of Georgia
HILTON, Thomas L., Carnegie Institute of Technology
HINDSMAN, Edwin, United States Office of Education
HOBERMAN, Solomon, New York City Department of Personnel
HOLLIS, Esther R., The Psychological Corporation
HONAKER, Linton R., Tuscarawas County Schools, Dover, Ohio
HOPKINS, Charles E., Atlanta (Georgia) Public Schools
HOPKINS, Florence M., Educational Testing Service
HOPKINS, George W., South Carolina State Department of Education
HOPMANN, Robert P., The Lutheran Church—Missouri Synod
HOROWITZ, George F., Columbia University
HORTON, Clark W., Dartmouth College
HOWELL, John J., Educational Testing Service
HUBBARD, John P., National Board of Medical Examiners
HUDDLESTON, Edith M., United States Office of Education
HUMPHRY, Betty J., Educational Testing Service
HUYSER, Robert J., Educational Testing Service
HYPPS, Irene C., District of Columbia Public Schools

IMPELLIZZERI, Irene H., Brooklyn College
IVEY, John E., Jr., Midwest Program on Airborne Television Instruction

JACKA, M. Elizabeth, Illinois State Scholarship Commission
JACKSON, Douglas N., The Pennsylvania State University
JACOBS, George, Detroit (Michigan) Public Schools
JACOBS, Paul, Educational Testing Service
JACOBSON, Audrey K., W. E. Howitt Junior High School, Farmingdale, New York

JANEBA, Hugo B., Rutherford (New Jersey) High School
JASPEN, Nathan, New York University
JOHNSON, A. Pemberton, Newark College of Engineering
JOHNSON, Janice M., United States Office of Education
JOHNSON, Richard A., United States Department of State

KABACK, Goldie Ruth, The College of the City of New York
KAMMAN, James F., University of Illinois
KATZ, Martin, Educational Testing Service
KAUFMANN, James D., Warwick Valley High School, Warwick, New York
KAVRUCK, Samuel, United States Office of Education
KEATING, Barry J., National Council of Churches
KELLEY, H. Paul, University of Texas
KENT, Clarence L., Virginia State Department of Education
KERSTING, Ethel, Educational Testing Service
KESSMAN, Maurice, Rochester Institute of Technology
KIMBALL, Elisabeth G., Educational Testing Service
KING, Mary L., Florence (South Carolina) Public Schools
KING, Richard G., Harvard University
KIP, Richard deR., American Institute for Property and Liability Underwriters
KIRKPATRICK, Forrest H., Wheeling Steel Corporation
KIRSCH, Arthur D., Datatrol Corporation
KLINE, William E., Baltimore County (Maryland) Board of Education
KNAPP, Thomas R., University of Rochester
KOCH, John C., Jr., River Dell Regional High School, Oradell, New Jersey
KOUTSOPOULOS, C. J., Educational Testing Service
KOZAK, Andrew V., Pennsylvania State University
KRALL, Walter R., Mamaroneck (New York) Public Schools
KRISTOF, Walter, Educational Testing Service
KUDER, G. Frederic, Duke University
KULKARNI, Sharad, Educational Testing Service

LA DUKE, Charles V., United States Armed Forces Institute
LAMBERT, Jane, Educational Testing Service
LANKTON, Robert S., Detroit (Michigan) Public Schools
LANNHOLM, Gerald V., Educational Testing Service
LAPOINTE, Archie, Educational Testing Service
LATHROP, Robert, University of Minnesota
LA VINE, Marian, Jericho (New York) High School
LAYTON, Wilbur L., Iowa State University
LEGGETT, Sidney, Glen Ridge (New Jersey) High School
LEIDNER, Burton R., Monmouth College
LENNON, Roger T., Harcourt, Brace and World
LEVINE, Harold G., Schenectady, New York
LEVINE, Richard, Educational Testing Service
LICHTENSTEIN, Ralph, New York City Department of Personnel
LIGON, Ernest M., Union College
LINDEMAN, Richard H., Columbia University
LINDQUIST, E. F., State University of Iowa
LITTERICK, William S., Keuka College
LORD, Frederic M., Educational Testing Service
LORET, Peter, Educational Testing Service
LORR, Maurice, Veterans Administration, Washington, D. C.

page 103

LOTTES, John J., College of Education at Geneseo
LOUGHLIN, Richard L., Bronx Community College
LOWERY, Zeb A., Rutherford County School Unit, Rutherfordton, North Carolina
LUCAS, Diana, Educational Testing Service
LUISETTE, Joan, New York City Board of Education
LUMSDAINE, A. A., University of California
LUTZ, Orpha M., Montclair State College

MACALUSO, Charles J., United States Naval Examining Center
MacGINITIE, Walter H., Columbia University
MADDOX, Clifford R., Cedarville College
MAGOON, Thomas, University of Maryland
MAIER, Milton, Educational Testing Service
MALCOLM, Donald J., Educational Testing Service
MALINCHOC, Joseph, Educational Testing Service
MALKIEL, Judith A., Educational Testing Service
MANLEY, Myron, Educational Testing Service
MANNELLO, George, Hofstra College
MAROHN, Mildred, The Lutheran Church—Missouri Synod
MARRON, Joseph E., United States Military Academy
MASON, Andrew W., Maryland State Department of Education
MASSEY, Will J., University of Maryland
MAYHEW, Lewis B., University of South Florida
McCANN, Forbes E., McCann Associates, Philadelphia, Pennsylvania
McCONVILLE, Carolyn, Educational Testing Service
McCORD, Richard B., City of Philadelphia Personnel Department
McCULLOUGH, C. William, Ohio State Department of Education
McDERMOTT, Robert F., United States Air Force Academy
McDILL, Mrs. Thomas, The Westminster Schools, Atlanta, Georgia
McINTIRE, Paul H., University of New Hampshire
McKENNA, Mae R., Crosby High School, Waterbury, Connecticut
McLAUGHLIN, Kenneth F., United States Office of Education
McMANUS, Leo F., Jr., J. P. Cleaver Company, Worcester, Massachusetts
McQUITTY, John V., University of Florida
McTARNAGHAN, Roy T., College of Education at Geneseo
McTERNAN, John W., College of Education at Plattsburgh
MEDLEY, Donald M., College of the City of New York
MELVILLE, S. Donald, Educational Testing Service
MERENDA, Peter F., University of Rhode Island
MERNYK, Charlotte Levy, Brooklyn, New York
MERRY, Robert W., Harvard University
MERWIN, Jack, University of Minnesota
MESSICK, Samuel, Educational Testing Service
MIDDLEMAS, Virginia, Center for Programed Instruction, New York City
MILES, Nelle H., Falls Church (Virginia) High School
MILHOLLAND, John E., University of Michigan
MILLER, Harold T., Houghton Mifflin Company
MILLETT, Esther, Westover School, Middlebury, Connecticut
MIRKIN, Louise, Educational Testing Service
MITCHELL, Blythe C., Harcourt, Brace and World
MITTMAN, Arthur, State University of Iowa

page 104

MITZEL, Harold E., College of the City of New York
MOLLENKOPF, William G., Procter and Gamble Company
MOORE, Zona G., Educational Testing Service
MOREAU, Edgar, New York State Department of Education
MORRIS, Darrell R., College Entrance Examination Board
MORRISON, Alexander W., Polytechnic Institute of Brooklyn
MOSER, Gene W., New York State Department of Education
MOSHER, Howard J., College of Education at New Paltz
MULLEN, Toni, Educational Testing Service
MUNSON, S. E., Franklin and Marshall College
MURRAY, John E., Office of Naval Research
MYERS, Charles T., Educational Testing Service
MYERS, Franklin G., Scarsdale (New York) High School
MYERS, Sheldon S., Educational Testing Service
MYERS, Isabel Briggs, Swarthmore, Pennsylvania

NOLAN, David M., Educational Testing Service
NOLL, Victor H., Michigan State University
NORMAN, Warren T., University of Michigan
NORRELL, Gwen, Michigan State University
NORTH, Robert D., Educational Records Bureau
NOSOW, Sigmund, Michigan State University

O'CONNOR, Virgil J., United States Air Force Academy
O'KEEFE, John T., Science Research Associates
OLIVER, R. A. C., Educational Testing Service
ORLEANS, Joseph B., George Washington High School, New York City
ORR, Cora I., Muskingum College
ORSHANSKY, Bernice, New Rochelle (New York) High School
ORTON, John, College of the City of New York
OTTOBRE, Frances M., Educational Testing Service

PACKARD, Albert G., Baltimore (Maryland) Public Schools
PALMER, Orville, Educational Testing Service
PARRES, JOHN G., Delaware State Department of Public Instruction
PASANELLA, Ann K., College Entrance Examination Board
PAYETTE, Roland, Educational Testing Service
PEEL, Edwin A., Birmingham (England) University
PERRY, William D., University of North Carolina
PETERSON, Donald A., Life Insurance Agency Management Association
PHILPOTT, Emily L., Queens College
PITCHER, Barbara, Educational Testing Service
PRESTON, B. Braxton, Educational Testing Service

QUINN, John S., Jr., Harcourt, Brace and World

RAIFFA, Gloria G., Pratt Institute
RAPHAEL, Brother Aloysius, Bishop Loughlin Memorial High School, Brooklyn, New York
RAPPARLIE, John H., Owens-Illinois Glass Company
READING, Ethel A., Educational Testing Service
REEHER, Kenneth R., Pennsylvania Department of Public Instruction

REID, John W., Cedar Crest College
REILLY, Pamela C., Educational Testing Service
REMMERS, H. H., Purdue University
RESSLER, Helene M., Educational Testing Service (Los Angeles Office)
RICKS, James H., Jr., The Psychological Corporation
RIDLEHOOVER, H. T., Scholastic Testing Service, Chicago, Illinois
RITENOUR, Louise R., Educational Testing Service
ROCA, Pablo, Pan American Union
ROCK, R. T., Harcourt, Brace and World
ROLLINGS, J. W., Colorado State Department of Education
ROSENGARTEN, William, Jr., Roslyn (New York) Public Schools
ROSENTHAL, Elsa, Educational Testing Service
ROSNER, Benjamin, Brooklyn College
ROSS, John A., Halifax (Nova Scotia) Department of Education
ROTH, Sidney, New York University
RUMMEL, J. Francis, University of Oregon
RUNKEL, Philip J., University of Illinois

SAIT, Edward, U. S. Army Enlisted Evaluation Center, Fort Harrison
SASAJIMA, Masu, College of the City of New York
SASLOW, Max S., New York City Department of Personnel
SCHAPIRO, Harold B., Young and Rubicam
SCHERER, Frank H., Rutgers, The State University
SCHUBERT, Herman J. P., Veterans Administration Hospital, Buffalo, New York
SCHWARZ, Mariette, Educational Testing Service
SCOTT, C. Winfield, Rutgers, The State University
SEIBEL, Dean W., Educational Testing Service
SERLING, Albert M., Educational Testing Service
SFORZA, Richard F., United States Military Academy
SHAHINIAN, Siroon, Services for the Aged, Flushing, New York
SHARP, Catherine G., Educational Testing Service
SHAYCOFT, Marion F., American Institute for Research
SHIELDS, Mary, National League for Nursing
SHIMBERG, Benjamin, Educational Testing Service
SIEGEL, Laurence, Miami University (Ohio)
SILVEY, Herbert M., State College of Iowa
SIMPSON, Elizabeth A., Illinois Institute of Technology
SINGLETON, Carlton M., Paterson State College
SITGREAVES, Rosedith, Columbia University
SKAGER, Rodney, Educational Testing Service
SKINNER, James A., Corpus Christi (Texas) Independent School District
SMITH, Alexander F., Southern Connecticut State College
SMITH, Allan B., University of Connecticut
SMITH, Ann Z., Educational Testing Service
SMITH, Lyman J., American Medical Association Honors Program
SMITH, Marshall P., Trenton State College
SMITH, Marshall S., Cambridge, Massachusetts
SNYDACKER, Susan, Educational Testing Service
SOHMER, Kenneth D., Millburn (New Jersey) High School
SOLOMON, Daniel, Center for Study of Liberal Education for Adults, Chicago, Illinois

SOLOMON, Robert J., Educational Testing Service
SOUTHER, Mary Tayloe, Tower Hill School, Wilmington, Delaware
SOUTHWORTH, J. Alfred, University of Massachusetts
SPANEY, Emma, Queens College
SPAULDING, Geraldine, Educational Records Bureau
SPEER, George S., Illinois Institute of Technology
SPENCER, Wesley G., Spencer Associates, Cambridge, Massachusetts
STAHL, Suzanne, Educational Testing Service
STAKE, Robert E., University of Nebraska
STECKLEIN, John E., University of Minnesota
STEFFEN, Hans, Prudential Insurance Company
STEINMAN, Arthur M., Trenton State College
STEWART, Mary, New York University Medical Center
STEWART, Naomi, Cranford, New Jersey
STICKELL, David W., Educational Testing Service
STILES, Grace Ellen, University of Rhode Island
STINCHFIELD, Philip A., West Hartford (Connecticut) Public Schools
STOKER, H. W., Florida State University
STONE, PAUL T., Huntingdon College, Montgomery, Alabama
STORM, Jerome F., Hofstra College
STOUGHTON, Robert W., Connecticut State Department of Education
STRICKER, Lawrence, Educational Testing Service
STUIT, Dewey B., State University of Iowa
STUNKARD, Clayton L., New York State Department of Education
SWANSON, Edward O., University of Minnesota
SWINEFORD, Frances, Educational Testing Service
TANENHAUS, Joseph, New York University

TANZER, Charles, LaSalle (New York City) Junior High School
TAYLOR, Justine, Educational Testing Service
TAYLOR, Marvin, Queens College
TERRANOVA, Christine, Columbia University
TERWILLIGER, James, Educational Testing Service
THOBURN, Isabella, Columbia College, Columbia, South Carolina
THOMPSON, Cary A., United States Air Force Academy
THOMPSON, Raymond E., Educational Testing Service
THORNDIKE, Robert L., Columbia University
THORNE, Margaret A., Educational Testing Service
THURSTONE, Thelma Gwinn, University of North Carolina
TIEDEMAN, David V., Harvard University
TOWNSEND, Agatha, Educational Records Bureau
TRAIL, Stanley M., Rhode Island College
TRAXLER, Arthur E., Educational Records Bureau
TREXLER, Laura M., Northern Valley Regional High School, Demarest, New Jersey
TRIGGS, Frances, Committee on Diagnostic Reading Tests, Mountain Home, North Carolina
TUCKER, Donald K., Northeastern University
TUCKER, Ledyard R, University of Illinois
TULLY, G. Emerson, Florida State University
TURNBULL, William W., Educational Testing Service
TURNER, Cornelius P., American Council on Education

page 107

URQUHART, Helen L., Brown University

VALLEY, John R., Educational Testing Service
VANCE, Edgar L., Air Force Institute of Technology
VICKERY, K. N., Clemson College
VICKERY, Verna, Bridgeport, Connecticut

WAGNER, E. Paul, Bloomsburg State College
WAGNER, Mazie Earle, College of Education at Buffalo
WAHLGREN, Hardy L., College of Education at Geneseo
WALLER, Raymond L., Allentown (Pennsylvania) School District
WALLMARK, Madeline M., Educational Testing Service
WALSH, John J., Boston College
WALTHER, Regis, U. S. Department of State
WALTON, Arline J., Malden (Massachusetts) Public Schools
WALTON, Wesley W., Educational Testing Service
WARRINGTON, Willard G., Michigan State University
WATKINS, Richard W., Educational Testing Service
WATSON, Walter S., The Cooper Union
WEAVER, J. Dale, Morristown (New Jersey) Public Schools
WEINER, Max, Brooklyn College
WEISS, Eleanor S., Educational Testing Service
WEISS, Joseph, Polytechnic Institute of Brooklyn
WEISS, Raymond A., New York University
WEITZ, Henry, Duke University
WELNA, Cecilia, University of Hartford
WESMAN, Alexander G., The Psychological Corporation
WEST, Elmer D., American Council on Education
WHITLA, Dean K., Harvard University
WILLARD, Richard W., Massachusetts Institute of Technology
WILLEY, Clarence F., Norwich University
WILLEY, Lawrence V., Jr., Science Research Associates
WILLIAMS, Clarence M., University of Rochester
WILLIAMS, Roger K., Morgan State College
WILLINGHAM, Warren W., Georgia Institute of Technology
WILLIS, William S., New York University
WILSON, Barbara S., Brooklyn College
WILSON, Kenneth M., Southern Regional Education Board, Atlanta, Georgia
WINANS, S. David, New Jersey State Department of Education
WINGO, Alfred L., Virginia State Department of Education
WINIEWICZ, Casimer S., U. S. Naval Examining Center
WINTERBOTTOM, John A., Educational Testing Service
WIRSIG, Jane, Educational Testing Service
WITHEY, Robert A., New Jersey State Department of Education
WOLINS, Leroy, Iowa State University
WOMER, Frank B., University of Michigan
WOOD, Dorothy Adkins, University of North Carolina
WOOD, James D., College of the City of New York
WRIGHT, Wilbur H., College of Education at Geneseo
WRIGHTSTONE, J. Wayne, New York City Board of Education

ZEIDNER, Joseph, Human Factors Research Branch, United States Army
ZEIGLER, Martin L., The Pennsylvania State University
ZOLA, Eugene J., Niskayuna High School, Schenectady, New York

page 108